

**Computer System Design and Operation in
the Railway and Other Transit Systems**

COMPUTERS IN RAILWAYS X



EDITORS

**J. Allen, C.A. Brebbia,
A.F. Rumsey, G. Sciutto,
S. Sone & C.J. Goodman**

***COMPUTERS IN
RAILWAYS X
COMPUTER SYSTEM DESIGN AND OPERATION IN THE
RAILWAY AND OTHER TRANSIT SYSTEMS***

WIT*PRESS*

WIT Press publishes leading books in Science and Technology.

Visit our website for the current list of titles.

www.witpress.com

WIT*eLibrary*

Home of the Transactions of the Wessex Institute.

Papers presented at COMPRAIL 2006 are archived in the WIT elibrary in volume 88 of

WIT Transactions on The Built Environment (ISSN 1743-3509).

The WIT electronic-library provides the international scientific community with immediate and permanent access to individual papers presented at WIT conferences.

<http://library.witpress.com>

TENTH INTERNATIONAL CONFERENCE ON
COMPUTERS IN RAILWAYS

COMPRAIL X

CONFERENCE CHAIRMEN

J. Allan

Rail Safety and Standards Board, UK

C. A. Brebbia

Wessex Institute of Technology, UK

A. F. Rumsey

Parsons Transportation Group

G. Sciutto

Universita di Genova

S. Sone

University of Kogakuin

C. J. Goodman

The University of Birmingham

INTERNATIONAL SCIENTIFIC ADVISORY COMMITTEE

T. Albrecht

W. Daamen

J. Esteves

I. A. Hansen

T. Koseki

N. Moreira

B. Ning

A. Radtke

J. Rodriguez

E. Schnieder

P. Tzieropoulos

A. Yoshimura

Sponsored by

WIT Transactions on the Built Environment

WIT Transactions on The Built Environment

Transactions Editor

Carlos Brebbia
Wessex Institute of Technology
Ashurst Lodge, Ashurst
Southampton SO40 7AA, UK
Email: carlos@wessex.ac.uk

Editorial Board

E Alarcon Universidad Politecnica de Madrid Spain	C Alessandri Universita di Ferrara Italy
S A Anagnostopoulos University of Patras Greece	E Angelino A.R.P.A. Lombardia Italy
H Antes Technische Universitat Braunschweig Germany	D Aubry Ecole Centrale de Paris France
D E Beskos University of Patras Greece	J J Bommer Imperial College London UK
F Butera Politecnico di Milano Italy	P G Carydis National Technical University of Athens Greece
J Chilton University of Nottingham UK	S Clement Transport System Centre Australia
M C Constantinou State University of New York at Buffalo USA	G Degrande Katholieke Universiteit Leuven Belgium
A De Naeyer Universiteit Ghent Belgium	W P De Wilde Vrije Universiteit Brussel Belgium
J Dominguez University of Seville Spain	F P Escrig University of Seville Spain
M N Fardis University of Patras Greece	C J Gantes National Technical University of Athens Greece
L Gaul Universitat Stuttgart Germany	Y Hayashi Nagoya University Japan
M Iguchi Science University of Tokyo Japan	L Int Panis VITO Expertisecentrum IMS Belgium
W Jager Technical University of Dresden Germany	C M Jefferson University of the West of England UK

- D L Karabalis**
University of Patras
Greece
- W Jager**
Technical University of Dresden
Germany
- W B Kratzig**
Ruhr Universitat Bochum
Germany
- J W S Longhurst**
University of the West of England,
UK
- L Lundqvist**
Unit for Transport and Location Analysis
Sweden
- G D Manolis**
Aristotle University of Thessaloniki
Greece
- F M Mazzolani**
University of Naples "Federico II"
Italy
- G Oliveto**
Università di Catania
Italy
- A S Papageorgiou**
Rensselaer Polytechnic Institute
USA
- A M Reinhorn**
State University of New York at Buffalo
USA
- C W Roeder**
University of Washington
USA
- M Saiidi**
University of Nevada-Reno
USA
- S A Savidis**
Technische Universität Berlin
Germany
- Q Shen**
Massachusetts Institute of Technology
USA
- P D Spanos**
Rice University
USA
- H Takemiya**
Okayama University
Japan
- E Taniguchi**
Kyoto University
Japan
- M A P Taylor**
University of South Australia
Australia
- E Kausel**
Massachusetts Institute of Technology
USA
- A N Kounadis**
National Technical University of Athens
Greece
- A A Liolios**
Democritus University of Thrace
Greece
- J E Luco**
University of California at San Diego
USA
- M Majowiecki**
University of Bologna
Italy
- G Matrisch**
DaimlerChrysler AG
Germany
- K Miura**
Kajima Corporation
Japan
- E Oñate**
Universitat Politècnica de Catalunya
Spain
- G G Penelis**
Aristotle University of Thessaloniki
Greece
- F Robuste**
Universitat Politècnica de Catalunya
Spain
- J M Roeset**
Texas A & M University
USA
- F J Sanchez-Sesma**
Instituto Mexicano del Petróleo
Mexico
- J J Sendra**
University of Seville
Spain
- A C Singhal**
Arizona State University
USA
- C C Spyarakos**
National Technical University of Athens
Greece
- I Takewaki**
Kyoto University
Japan
- J L Tassoulas**
University of Texas at Austin
USA
- R Tremblay**
Ecole Polytechnique
Canada

R van der Heijden

Radboud University
Netherlands

A Yeh

The University of Hong Kong
China

R Zarnic

University of Ljubljana
Slovenia

R van Duin

Delft University of Technology
Netherlands

M Zador

Technical University of Budapest
Hungary

This page intentionally left blank

***COMPUTERS IN
RAILWAYS X
COMPUTER SYSTEM DESIGN AND OPERATION IN
THE RAILWAY AND OTHER TRANSIT SYSTEMS***

Editors

J. Allan

Rail Safety and Standards Board, UK

C. A. Brebbia

Wessex Institute of Technology, UK

A. F. Rumsey

Parsons Transportation Group

G. Sciutto

Universita di Genova

S. Sone

University of Kogakuin

C. J. Goodman

The University of Birmingham

WITPRESS Southamton, Boston



J. Allan

Rail Safety and Standards Board, UK

C. A. Brebbia

Wessex Institute of Technology, UK

A. F. Rumsey

Parsons Transportation Group

G. Sciutto

Universita di Genova

S. Sone

University of Kogakuin

C. J. Goodman

The University of Birmingham

Published by

WIT Press

Ashurst Lodge, Ashurst, Southampton, SO40 7AA, UK

Tel: 44 (0) 238 029 3223; Fax: 44 (0) 238 029 2853

E-Mail: witpress@witpress.com

<http://www.witpress.com>

For USA, Canada and Mexico

Computational Mechanics Inc

25 Bridge Street, Billerica, MA 01821, USA

Tel: 978 667 5841; Fax: 978 667 7582

E-Mail: infousa@witpress.com

<http://www.witpress.com>

British Library Cataloguing-in-Publication Data

A Catalogue record for this book is available
from the British Library

ISBN: 1-84564-177-9

ISSN: 1746-4498

The texts of the papers in this volume were set individually by the authors or under their supervision. Only minor corrections to the text may have been carried out by the publisher.

No responsibility is assumed by the Publisher, the Editors and Authors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

© WIT Press 2006.

Printed in Great Britain by Cambridge Printing.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the Publisher.

Preface

COMPRAIL 2006, the tenth in a series of well-established and successful international conferences on Computer System Design and Operation in the Railway and Other Transit Systems, was held in Prague, Czech Republic, in 2006. Since 1987, COMPRAIL has provided a world forum for planners, designers, manufacturers and operators to discuss how they can benefit from computer-based techniques.

This book contains most of the papers presented at COMPRAIL 2006, representing the latest research, from development and application of computers to the management, design, manufacture and operations of railways and other passenger, freight and transit systems.

The Conference attracted a large number of papers, divided into the following sections: Planning; Safety; Passenger interface systems; Decision support systems; Computer techniques; Converting metros to driverless operation; Advanced train control; Train location; Dynamic train regulation; Timetable planning; Operations quality; Communications; Energy management; Power supply; Dynamics and wheel/rail interface; Freight; Condition monitoring.

This book is distributed throughout the world by WIT Press, the publishing arm of the Wessex Institute of Technology. In addition, together with all other COMPRAIL Conferences held from 1994 onwards, the papers are displayed in the electronic library of the Transactions of the Wessex Institute, where they are permanently available to the international scientific community.

The Editors are grateful to all the authors for the excellent papers and to those members of the International Scientific Advisory Committee who participated in the review process. The success of the conference and this book is the result of their significant contribution of time and energy.

The Editors
Prague, 2006

This page intentionally left blank

Contents

Section 1: Planning

Choices between stairs, escalators and ramps in stations <i>W. Daamen, P. H. L. Bovy & S. P. Hoogendoorn</i>	3
How is the business case used by stakeholders for making project decisions with PFI/PPP projects? <i>M. J. Gannon</i>	13
A bilevel model for optimizing station locations along a rail transit line <i>S. Samanta & M. K. Jha</i>	23
Design and implementation of virtual environments for planning and building sustainable railway transit systems <i>M. Chandramouli, B. Huang, T.-Y. Chou, L. K. Chung & Q. Wu</i>	33
The network effects of railway investments <i>S. Hansen, A. Landex & A. H. Kaas</i>	45
Modelling and simulation of the traffic management in a migration phase: example of “Ligne 1” of the Parisian subway <i>M. Ghantous-Mouawad, W. Schön, J.-L. Boulanger & G. Churchill</i>	55
Computer based ex-ante evaluation of the planned railway line between Copenhagen and Ringsted by use of a Decision Support System named COSIMA-DSS <i>K. B. Salling & A. Landex</i>	65
Assessing rail transport network performance and reliability <i>R. Raicu & M. A. P. Taylor</i>	75
A study on a mathematical model of the track maintenance scheduling problem <i>S. M. Oh, J. H. Lee, B. H. Park, H. U. Lee & S. H. Hong</i>	85

Statistical method for the evaluation of railway system modifications <i>M. Chandesris</i>	97

Prioritized Rail Corridor Asset Management <i>T. Selig & J. Kasper</i>	105

Section 2: Safety

An assessment of hazard probability due to Pentium processor errata in automatic train control applications <i>C. Bantin</i>	115

Role of supervision systems in railway safety <i>F. Belmonte, J.-L. Boulanger, W. Schön & K. Berkani</i>	129

Automatic train controller safety simulation <i>R. A. V. Gimenes, J. R. de Almeida Jr. & T. R. Nogueira</i>	139

Common approach for supervising the railway safety performance <i>E. M. El Koursi & L. Tordai</i>	147

Potential dangerous object detection on railway ballast using digital image processing <i>P. L. Mazzeo, E. Stella, M. Nitti & A. Distante</i>	157

Thermal characteristics of novel brake friction materials for light rail transit applications <i>N. Valliyappan, D. Berhan, M. N. Darius & G. Solomon</i>	167

Section 3: Passenger interface systems

A fast method for estimating railway passenger flow <i>Y. Nagasaki, M. Asuka & K. Komaya</i>	179

Route-choice support system for passengers in the face of unexpected disturbance of train operations <i>R. Tsuchiya, Y. Sugiyama, K. Yamauchi, K. Fujinami, R. Arisawa & T. Nakagawa</i>	189

A new delay forecasting system for the Passenger Information Control system (PIC) of the Tokaido-Sanyo Shinkansen <i>K. Fukami, H. Yamamoto, T. Hatanaka & T. Terada</i>	199

Section 4: Decision support systems

Reconstruction of train trajectories from track occupation data to determine the effects of a Driver Information System <i>T. Albrecht, R. M. P. Goverde, V. A. Weeda & J. van Luipen</i>	207
A decision support system for track maintenance <i>C. Meier-Hirmer, A. Senée, G. Riboulet, F. Sourget & M. Roussignol</i>	217
The new Shinkansen rescheduling system for drivers and crew <i>H. Shimizu, H. Tanabe, S. Honda & K. Yasura</i>	227
A Decision Support System for railway timetabling (MOM): the Spanish case <i>F. Barber, P. Tormos, A. Lova, L. Ingolotti, M. A. Salido & M. Abril</i>	235
Measurement of train driver's brain activity by functional near-infrared spectroscopy (fNIRS) <i>T. Kojima, H. Tsunashima & T. Y. Shiozawa</i>	245
Methodology for the LCC-Analysis and the optimal migration of the railway operations control on the example of ETCS <i>M. Obrenovic, B. Jaeger & K. Lemmer</i>	255
Application of location detection system using active type RFID tags to railways <i>K. Seki, S. Suzuki, M. Ukai & R. Tsuchiya</i>	265
A survey on SNCF decision support system tools to supervise and to pilot train traffic in operation <i>D. Gauyacq</i>	275

Section 5: Computer techniques

Distributed constraint satisfaction problems to model railway scheduling problems <i>P. Tormos, M. Abril, M. A. Salido, F. Barber, L. Ingolotti & A. Lova</i>	289
Blocking time reduction for level crossings using the genetic algorithm <i>Y. Noguchi, H. Mochizuki, S. Takahashi, H. Nakamura, S. Kaneko & M. Sakai</i>	299

Modeling a distributed railway interlocking system with object-oriented Petri nets <i>X. Hei, H. Mochizuki, S. Takahashi, H. Nakamura, M. Fukuda, K. Iwata & K. Sato</i>	309

Using UML diagrams for system safety and security environment analysis <i>F. M. Rachel & P. S. Cugnasca</i>	319

System-independent and quality tested availability of railway data across country and system borders by the model driven approach <i>H. R. Gnägi & N. Stahel</i>	329

Formalisation and simulation of operating rules using coloured Petri nets <i>O. Lahlou, P. Bon & L. Allain</i>	341

From UML to B – a level crossing case study <i>J.-L. Boulanger, P. Bon & G. Mariano</i>	351

Section 6: Converting metros to driverless operation (special section organised by A. F. Rumsey)

The feasibility case for converting existing heavy metro systems to driverless operation <i>P. Thomas</i>	363

Converting existing service to fully automatic operation <i>F. Fabbian</i>	373

Re-signaling the Paris Line 1: from driver based to driverless operation <i>C. Braban & P. Charon</i>	381

Application of communication based Moving Block systems on existing metro lines <i>L. Lindqvist & R. Jadhav</i>	391

Driverless CBTC – specific requirements for CBTC systems to overcome operation challenges <i>M. P. Georgescu</i>	401

Section 7: Advanced train control

CBTC (Communication Based Train Control): system and development <i>N. Bin, T. Tao, Q. K. Min & G. C. Hai</i>	413

An algorithm for braking curve calculations in ERTMS train protection systems <i>B. Friman</i>	421
ICONIS: the window for URBALIS controlled automatic METRO <i>P. Noury</i>	431
Automatic train operation system for the high speed Shinkansen train <i>Y. Yasui</i>	441

Section 8: Train location

Design of experiment for the validation of ATP/ATC odometry algorithms <i>M. Malvezzi, G. Cocci & A. Tarasconi</i>	449
Location in railway traffic: generation of a digital map for secure applications <i>F. Böhringer & A. Geistler</i>	459
Train position detecting system using radio millimeter-waves <i>T. Maeda, K. Watanabe & M. Ono</i>	469

Section 9: Dynamic train regulation

Optimal train control at a junction in the main line rail network using a new object-oriented signalling system model <i>R. Takagi, P. F. Weston, C. J. Goodman, C. Bouch, J. Armstrong, J. Preston & S. Sone</i>	479
Optimising train priorities to support the regulation of train services with the assistance of active and deductive databases <i>C. Sakowitz & E. Wendler</i>	489
Simulation of traffic management with FRISO <i>A. D. Middelkoop & L. Loeve</i>	501
Influences of station length and inter-station distance on delays and delay propagation on single-track lines with regional rail traffic <i>O. Lindfeldt</i>	511
Simulation of disturbances and modelling of expected train passenger delays <i>A. Landex & O. A. Nielsen</i>	521

Running time re-optimization during real-time timetable perturbations <i>A. D'Ariano & T. Albrecht</i>	531

ALFa – a software tool for optimal scheduling of demand oriented train services <i>S. Scholz & T. Albrecht</i>	541

An algorithm for train rescheduling using rescheduling pattern description language R <i>C. Hirai, N. Tomii, Y. Tashiro, S. Kondou & A. Fujimori</i>	551

Section 10: Timetable planning

State-of-the-art of railway operations research <i>I. A. Hansen</i>	565

Timetable management and operational simulation: methodology and perspectives <i>A. Radtke</i>	579

The contribution of state resources in a constraint-based scheduling model for conflict solving at railway junctions <i>J. Rodriguez</i>	591

A new idea for train scheduling using ant colony optimization <i>K. Ghoseiri</i>	601

Joint design standard for running times, dwell times and headway times <i>V. A. Weeda & P. B. L. Wiggendaad</i>	611

RTCSIM: an innovative, extendable computation engine for timetable validation <i>T. Polzin & R. Goßmann</i>	621

Evaluating stochastic train process time distribution models on the basis of empirical detection data <i>J. Yuan, R. M. P. Goverde & I. A. Hansen</i>	631

Section 11: Operations quality

Practical use of the UIC 406 capacity leaflet by including timetable tools in the investigations <i>A. Landex, A. H. Kaas, B. Schittenhelm & J. Schneider-Tilli</i>	643

A method to estimate passenger flow with stored data at ticket gates <i>S. Myojo</i>	653

Analysis and optimisation of railway nodes using simulation techniques <i>A. Kavička, V. Klima & N. Adamko</i>	663

Section 12: Communications

Model checker for railway signalling communication protocol <i>J.-G. Hwang, H.-J. Jo & J. H. Lee</i>	675

A new ground-to-train communication system using free-space optics technology <i>H. Kotake, T. Matsuzawa, A. Shimura, S. Haruyama & M. Nakagawa</i>	683

Communications security concerns in communications based train control <i>M. Hartong, R. Goel & D. Wijesekera</i>	693

Ethernet-based network with high reliability for railway signaling systems <i>H.-J. Jo, J.-G. Hwang & Y.-K. Yoon</i>	703

Section 13: Energy management

Information system for railway energy management <i>G. Hribar, B. Dremelj & M. Sekavčnik</i>	713

Impact of train model variables on simulated energy usage and journey time <i>P. Lukaszewicz</i>	723

Section 14: Power supply

Investigation into the computational techniques of power system modelling for a DC railway <i>A. Finlayson, C. J. Goodman & R. D. White</i>	735

Catenary and autotransformer coupled optimization for 2x25kV systems planning <i>E. Pilo, L. Rouco & A. Fernandez</i>	747

A study of capacity calculation of regenerative inverter for 1500V DC traction system <i>C. H. Bae, M. S. Han, Y. K. Kim, S. Y. Kwon & H. J. Park</i>	757

Train operation minimizing energy consumption in DC electric railway with on-board energy storage device <i>K. Matsuda, H. Ko & M. Miyatake</i>	767
Calculations and measurements of harmonic current distributions in the catenary of railways with single-phase A.C. <i>A. Zynovchenko, J. Xie, S. Jank & F. Klier</i>	777
A numerical algorithm for run-curve optimization of trains considering a DC feeding circuit <i>H. Ko & M. Miyatake</i>	787
Railway modelling for power quality analysis <i>M. Chymera, A. C. Renfrew & M. Barnes</i>	797
A mixed AC/DC model for railway power systems <i>J. Muñoz-Riesco, E. Pilo, A. Fernandez & P. Cucala</i>	805
A user interface for the representation of the dynamic results on the pantograph-catenary interactions <i>D. Cebrián, T. Rojo, A. Alberto, E. Arias, F. Cuartero & J. Benet</i>	817

Section 15: Dynamics and wheel/rail interface

Dynamic identification of a 1:5 scaled railway bogie on roller rig <i>N. Bosso, A. Gugliotta & A. Somà</i>	829
On enhanced tilt strategies for tilting trains <i>B. Kufver & R. Persson</i>	839
Railway car dynamic response to track transition curve and single standard turnout <i>J. Drożdżel & B. Sowiński</i>	849
Optimization of special freight wagons with small wheel diameter <i>A. Rindi, D. Fioravanti, L. Pugi, M. Rinchi & J. Auciello</i>	859
Design and simulation of railway vehicles braking operation using a scaled roller-rig <i>N. Bosso, A. Gugliotta & A. Somà</i>	869
Study on vertical dynamic vehicle-track interactions using the TRADYS test facility and computer simulation <i>M. Miwa & A. Yoshimura</i>	885

The study of design on the urban track transportation vibration prevention: floating slab track in the bored tunnel at Taipei <i>Sy. Chang, K. Y. Chang, K. H. Cheng & R. J. T. Chen</i>	895

Section 16: Freight

Effect of the distribution of the arrivals and of the intermodal unit sizes on the transit time through freight terminals <i>G. Malavasi, A. Quattrini & S. Ricci</i>	905
COMPAT: a decision support tool for determining the necessity of rail infrastructure <i>M. J. Wolbers</i>	915
A cost effective solution to manage rail cargo fleets: the final assessment of the F-MAN project <i>G. Cosulich, A. Derito, M. Giannettoni & S. Savio</i>	923

Section 17: Condition monitoring

Methodology for the monitoring, control and warning of defects for preventive maintenance of rails <i>K. C. Labropoulos, P. Moundoulas & A. Moropoulou</i>	935
Onboard measurement method for signaling equipment on Probe Trains <i>H. Nakamura, S. Takahashi, T. Hiramoto, H. Mochizuki & T. Mizuma</i>	945
Fault detection of railway track by multi-resolution analysis <i>T. Kojima, H. Tsunashima & A. Matsumoto</i>	955
Research on the onboard auto-test system for track circuit compensating capacitors <i>H. Zhao, Y. Liu & B. Liu</i>	965
Monitoring wheel defects on a metro line: system description, analysis and results <i>M. Seco, E. Sanchez & J. Vinolas</i>	973
Author index	983

This page intentionally left blank

Section 1

Planning

This page intentionally left blank

Choices between stairs, escalators and ramps in stations

W. Daamen, P. H. L. Bovy & S. P. Hoogendoorn

*Department of Transport & Planning, Delft University of Technology,
The Netherlands*

Abstract

In assessing the design of a public transfer station it is very important to be able to predict the routes taken by the passengers. Most simulation tools use very simple route choice models, only taking into account the shortest walking distance or walking time between a passenger's origin and destination. In order to improve this type of route choice model, other factors affecting passenger route choice need to be identified. Also, the way these factors influence route choice behaviour needs to be determined, indicating how each factor is valued. In this research, route choice data have been collected in two Dutch train stations by following passengers through the facility from their origins to their destinations. These data have been used to estimate extended route choice models. The focus in this contribution is on the influences of level changes in walking routes on passenger route choice behaviour. It appears that the different ways of bridging level changes (ramps, stairs, escalators) each have a significant and different impact on the attractiveness of a route to the traveller.

Keywords: passenger behaviour, route choice, vertical infrastructure, choice models.

1 Introduction

Public transport passengers are subjected to the problem of making route choice decisions in interchange nodes. This route choice becomes more and more complex, since public transport stations are slowly turned into multi-purpose facilities, including shopping and catering services. While this increases the number of activities that may be performed in a station, it also makes the station



less understandable and thus complicates the route choice process for individuals.

In assessing public transport facility design, understanding pedestrian route choice is very important, since it is one of the key factors affecting the occurrence of crowding. In illustration: when all pedestrians take the same route, large concentrations of pedestrians will occur, whereas when interchanging passengers are more evenly distributed over the different routes in the station, the infrastructure will be more efficiently used and pedestrian comfort – expressed in terms of densities and speeds [1] – is likely to be higher.

Unfortunately, most simulation tools used for such assessments only include a fairly simple route choice model, derived from shortest path models. We argue that to improve our knowledge on the route choice process and to identify the factors affecting this process more varied and detailed observations are needed. With such observations, we may be able to derive more complex and accurate route choice prediction models.

This paper presents results from an effort to improve the predictive performance of pedestrian route models for public transport facilities by collecting dedicated data and estimating route choice models. We focus on the influences of level changes in walking routes and the different impacts of various ways of bridging these changes of level such as by ramps, escalators, and stairs.

This contribution starts with an overview of the state-of-the-art on route choice in public transport facilities. The main part of the contribution consists of the data collection and the estimation of route choice models including various variables. We end with some conclusions.

2 Passenger route choice in public transport nodes

For a general reference about travel behaviour in transfer stations, see [2]. This paper is on the influence of level changes on passenger route choice in public transport facilities. Reason for the hypothesized large influence is the fact that bridging level differences costs physical effort to the pedestrian. Physical abilities have major importance in the route choice process. Personal characteristics are important in order to see their effects on route choice.

Considering the vertical dimension, pedestrians do not only consider walking time (including delays in front of and on escalators and stairs) and walking distance, but also the effort involved in climbing a grade (with similar travel times, only few pedestrians choose the stairs in ascending direction) [3].

Especially in transfer stations, walking in vertical direction is very important. Apart from Cheung and Lam's study [3] no systematic analyses of the route choice impact of different forms of bridging level changes are known. Therefore, the study described in this paper focuses on the influence of level changes in routes and the type of their overcoming on passenger route choice behaviour. To that end, we have tried to gather a more extensive route data set, also taking into account several other factors identified in literature as potentially affecting route choice behaviour.



3 Adopted research approach

The research question to be dealt with in this paper may be formulated as: ‘Do the various types of walking facility influence passenger’s route choice in train stations, with special attention to the way of vertical height bridging?’ We try to address this question by empirical observations of revealed choices of train users under varying conditions with respect to trip lengths, origins and destinations, choice set compositions and route choice factors, especially types of vertical height bridging (level, stairs, escalators, ramps). To that end, passengers’ route choices will be observed in different railway stations having enough variability in route choice conditions. The observations will be used to derive parameter estimates by adopting discrete choice models.

To estimate whether height bridging facilities affect route choice in stations we choose to adopt a path-size logit model since this model type is known to be sufficiently robust to cope with the necessary simplifying assumptions. In this paper, we will focus on the data collection and the results of the choice estimation. For more details on the choice estimation, we refer to [4].

4 Passenger route data collection in stations

New data had to be collected in order to enable addressing the research question on level difference impacts on passenger route choice in public transport facilities. We chose the Dutch railway stations of the cities of Delft and Breda (the Netherlands) for this data collection, since the number of different choice situations and route alternatives is reasonably high in these stations. Also, the total range of infrastructure elements, that is, level elements, stairs, escalators, and ramps, can be found by pooling the data of Delft and Breda. The average daily volume in both stations is similar (21,750 passengers in Breda and 19,500 passengers in Delft). Also, the distribution over the trip motives is similar (38% work and business, 27% school and study and 36% social and leisure). Differences in route choice are therefore mainly affected by the infrastructure layout.

The data collection on chosen routes has been performed by unobtrusively following passengers, while monitoring personal characteristics (gender, age) as well as the route they chose from an entry point to the exit point.

In Delft the choice situations to be observed differ in various respects such as in length, in length of the sheltered part of the route, in level difference, and in facility type for bridging these heights (stairs or ramp). Only trips to and from platform 1 (see Figure 1:) are taken into account in this study, since no route alternatives exist to platform 2.

In Breda, alternative routes differ in the type of facility for bridging levels (stairs or escalator). Figure 1: shows plans of both stations.

The observations in Delft have taken place during both the morning peak hour and the evening peak hour during three weeks in November 2003. In Breda, observations have also been performed during both peak hours, but only during one week in January 2004.



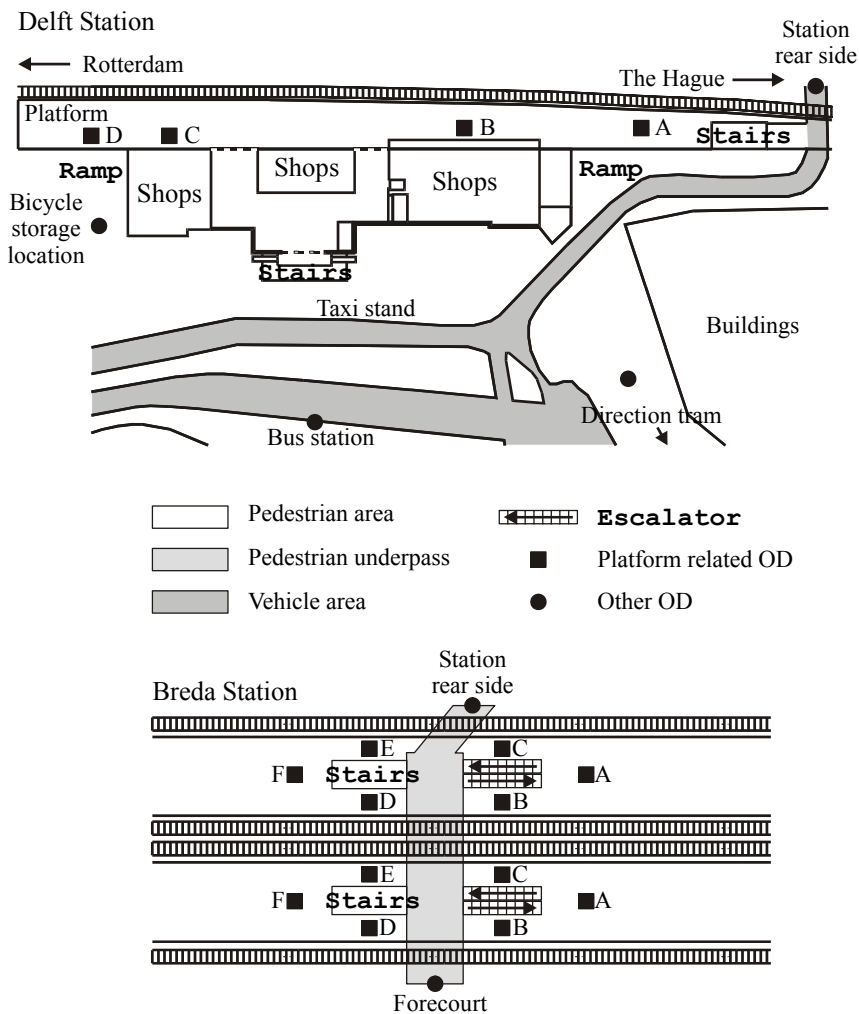


Figure 1: Plans of Delft Station and Breda Station (below).

Fixed characteristics of the stations have been observed once during an off-peak period. All other observations have been made while following the passengers from their origin, using a stopwatch to record walking times and writing the data on a dedicatedly designed observation form.

In Delft, 745 observations have been collected for 68 different origin-destination combinations, whereas in Breda 180 observations have been collected for 48 origin-destination combinations. The level changes in Delft are limited and are bridged by ramps and (short) stairs, whereas in Breda pedestrians have to choose between stairs and escalators to arrive at the platform. The average number of routes to choose between is larger in Delft (3-4 route alternatives) than in Breda (2-3 routes).

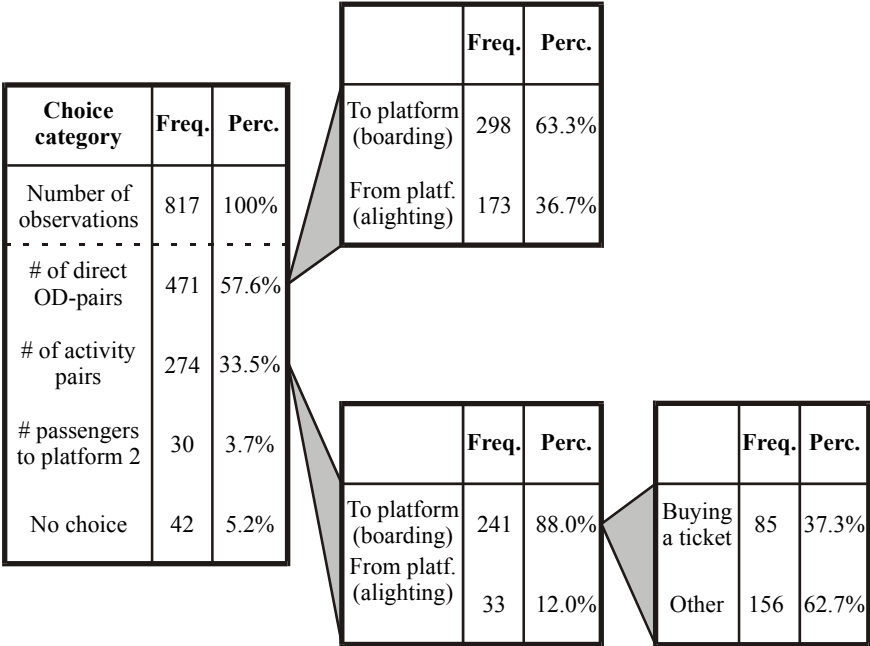


Figure 2: Overview of trip types observed in Delft Station.

Figure 2: gives an overview of various observed trip types in Delft according to direction and activity participation in the station. Trips with intermediate activities are split into separate direct routes. Trips to and from platforms are assumed to show different route choice behaviour. Table 1 summarises some of the observed person characteristics in both Delft Station and Breda Station. Unfortunately, only few observations with heavy luggage are available in the sample.

In order to give the reader a feeling of the collected data, Table 2 and Table 3 show some results for station Breda. Table 2 shows the route observations pertaining to the trips in Breda between platform 5/6 and the station hall. For each of the 16 distinguished origin-destination relations the number of route alternatives (used and non-used) and some attributes of chosen routes are given. The table shows sufficient variability in walking times between relations as well as in composition of stairs and escalators. It shows that only in a minority of cases all alternatives are indeed used by travellers.

Table 3 summarises for each route in a choice situation the number of pedestrians having chosen this route as well as the corresponding average walking time. Only the trips in Breda Station between platform 5/6 and the hall are included in the table. The routes in this table are numbered according to the location where they enter the platform. Route 1 is via the escalator on the right, while route 2 uses the stairs on the left. Route 3 is the route using the stairs



Table 1: Overview of observed characteristics.

	Delft		Breda		Total	
	Freq.	Perc.	Freq.	Perc.	Freq.	Perc.
Men	452	60.7%	93	51.7%	545	58.9%
Women	293	39.3%	87	48.3%	380	41.1%
Children / students	218	29.3%	63	35.0%	281	30.4%
Commuters	509	68.3%	115	63.9%	624	67.5%
Seniors	18	2.4%	2	1.1%	20	2.2%
With luggage	25	3.4%	11	6.1%	36	3.9%
Without luggage	720	96.6%	169	93.9%	889	96.1%

Table 2: Some characteristics for route choices in Breda Station.

Origin-destination relation	# Alternatives		Min. travel time [s]	Mean travel time [s]	St. dev. Travel time [s]	# Observations	# Alternatives with			
	#	Used alternatives					Stairs up	Stairs down	Escalator up	Escalator down
Station hall – platform 5/6 part A	3	1	15	51,7	13,9	35	2	0	1	0
Station hall – platform 5/6 part B	2	2	50	55	7,1	2	1	0	1	0
Station hall – platform 5/6 part C	2	1	56	62	6,6	3	1	0	1	0
Station hall – platform 5/6 part D	2	1	48	48	-	1	1	0	1	0
Station hall – platform 5/6 part E	2	1	28	28	-	1	1	0	1	0
Station hall – platform 5/6 part F	3	2	32	47,6	8,9	14	1	0	2	0
Platform 5/6 part A – station hall	3	1	45	51	8,5	2	0	2	0	1
Platform 5/6 part B – station hall	2	1	62	72	14	3	0	1	0	1
Platform 5/6 part C – station hall	2	-	-	-	-	0	0	1	0	1
Platform 5/6 part D – station hall	2	1	38	48	14	3	0	1	0	1
Platform 5/6 part E – station hall	2	-	-	-	-	0	0	1	0	1
Platform 5/6 part F – station hall	3	1	49	50	1,4	2	0	1	0	2



or A, route 1 includes the escalator, route 2 includes the stairs and parts C and E of the platform, while route 3 also includes the stairs as well as parts B and D of the platform). Sometimes, the number of routes is limited, since one of the routes is worse in all aspects. This route is then left out.

Table 3: Observed route choices in Breda Station.

Origin-destination relation	# Observations	Route 1		Route 2		Route 3	
		# Chosen	Mean travel time	# Chosen	Mean travel time	# Chosen	Mean travel time
Station hall – platform 5/6 part A	35	35	51.7	0	79.3	0	79.3
Station hall – platform 5/6 part B	2	1	52.0	1	60.7		
Station hall – platform 5/6 part C	3	3	62.0	0	74.9		
Station hall – platform 5/6 part D	1	0	63.5	1	48.0		
Station hall – platform 5/6 part E	1	0	37.0	1	28.0		
Station hall – platform 5/6 part F	14	0	72.4	13	45.7	1	72.4
Platform 5/6 part A – station hall	2	2	51.0	0	78.8	0	78.8
Platform 5/6 part B – station hall	3	3	72.0	0	89.9		
Platform 5/6 part C – station hall	0	-	-	-	-		
Platform 5/6 part D – station hall	3	0	67.4	3	48.0		
Platform 5/6 part E – station hall	0	-	-	-	-		
Platform 5/6 part F – station hall	2	0	80.9	2	50.0	0	80.9

925 out of 1010 observations appeared useful for model estimation. The data clearly show that travel time is an important factor in route choice. Table 3 shows that many passengers choose for the route with the shortest travel time. The data collected in Breda also shows that passengers prefer the use of the escalator instead of the stairs. We will elaborate on these findings more specifically using choice model estimation in section 5.

5 Model estimation outcomes

We will confine ourselves to a discussion of the outcomes of the model estimation and refer to [4] and [5] for details about the estimation procedure.

First of all, we can conclude (as expected) that routes with short walking times are preferred over routes with longer walking times. This is a general observation, corresponding to other findings in literature, and independent of the types of facilities routes consist of.

Secondly, the facility types in routes do influence the route choice behaviour of passengers. We distinguish between level elements (platform, corridors and hallways), stairs, escalators, and ramps. Passengers prefer walking on a level



element, since this is the most natural way of walking. However, in most stations tracks are crossed at different levels for safety reasons. Passengers are therefore obliged to overcome differences in height using stairs, escalators, or ramps. Comparing the different facility types, stairs are valued much more negative than escalators and ramps. While escalators and ramps are valued similar, escalators are slightly preferred.

For a quantitative comparison, we use the relative level penalty, comparing walking time on a specific type of facility to walking time on a level element. Table 4 shows the results for such a comparison. We see that walking time on stairs is valued 1.86 times that of walking on a level element, thus 1 second walking on stairs is valued similar to 1.86 seconds walking on a level element. For escalators and ramps, a single second of walking time is equal to respectively 1.28 seconds and 1.37 seconds of walking on a level element.

When we distinguish between the directions a facility is used (upwards versus downwards) we see that passengers prefer the downward direction (lower relative level penalties). Walking downwards on an escalator or ramp is even preferred over walking on level elements.

Table 4: Relative level penalties for different types of facilities.

	All directions	Upwards	Downwards
Level element	<i>l</i>	<i>l</i>	
Stairs	1.86	2.78	n.s.
Escalator	1.28	1.86	0.72
Ramp	1.37	1.66	0.95

n.s. = non-significant

In addition to the influence of the facility type, we have also looked at the influence of passenger characteristics, trip characteristics and observational characteristics, see [4] and Table 5. In Table 5, the second column indicates the reference group (in italics). The weight that the other groups give to travel time is given relative to this reference group.

With respect to passenger characteristics, we tested differences in choice behaviour between men and women, between children, adults, and seniors, and between passengers with and without luggage. The only considerable distinction in behaviour is found for seniors, having a very strong preference for short routes (nearly eight times as much as adults). The trip characteristics (boarding or alighting passengers and whether passengers performed an intermediate activity or not) did not change passenger’s behaviour.

Finally, observational characteristics have been looked at. No differences were found in route choice during the morning peak and the evening peak. However, weather conditions and the day of the week did influence route choice behaviour. Bad weather conditions made passengers choose even more for the shortest routes, whereas passengers valued walking times significantly worse on Monday than on Tuesday and Thursday. This might also be due to the passenger type (less regular travellers on Monday versus regular travellers on the other days of the week).



Table 5: Influence of passenger, trip and observational characteristics.

Personal characteristics	Reference group			
	Men	Women		
	<i>I</i>	1.00		
	Children	Adults	Seniors	
	<i>I</i>	0.82	6.34	
	With luggage	Without luggage		
	<i>I</i>	0.92		
Trip characteristics	Boarding	Alighting		
	<i>I</i>	1.21		
	With activity	Without activity		
	<i>I</i>	1.09		
Observational characteristics	Morning peak	Evening peak		
	<i>I</i>	0.93		
	Sunny weather	Clouded		
	<i>I</i>	1.58		
	Monday	Tuesday	Wednesday	Thursday
	<i>I</i>	0.71	0.79	0.69

6 Conclusions

In this paper, we have discussed data collection for modelling passenger route choice in public transfer stations. These data have been collected in two Dutch train stations by following passengers and recording their chosen route as well as some personal characteristics. Based on these data a route choice model has been estimated, taking into account not only walking times, but also the influences of different types of facilities (level elements, stairs, escalators, and ramps). The estimated model came up with very reliable results. Including a route choice model like this with such explanatory power in a passenger flow simulation tool, the assignment of passengers to the station infrastructure network will be more accurate, thus leading to better predictions of the simulation tool with respect to concentrations of passengers in a station.

Although intended for use in a simulation tool, the values for infrastructure types estimated in this research may directly be applied in facility design decisions on supporting e.g. the installation cost of escalators. The estimated values are passengers' qualifications of the infrastructure elements, different for upward and downward direction and as such an indication for the distribution of the passengers over the elements. Cost and comfort of an infrastructure type may be assessed directly in the design stage, together with the offered capacity.

Although this is already a remarkable improvement, further research is recommended. First, the existing data set might be applied for more detailed models. More variables, also with respect to personal characteristics and observational characteristics will be included in a single utility function. Also,



other types models will be looked at. To be able to predict passenger route choice in an arbitrary station or already in the design stage, more data will be collected on other stations with different layouts. We will look for a more extended number of infrastructure types (elevators), where also characteristics of the infrastructure will differ (length, grade). We also plan to observe different passenger types, focussing among other things on the amount of luggage that passengers carry, passengers' habit and passenger behaviour in groups or travelling with small children.

Acknowledgement

This publication has been produced in the framework of the Seamless Multimodal Mobility research program, which is carried out by the Netherlands TRAIL Research School for Transport, Infrastructure and Logistics.

References

- [1] Transportation Research Board, *Highway Capacity Manual*, National Academy of Sciences: Washington D.C., 2000.
- [2] Daamen, W. *Modelling passenger flows in public transport facilities*, PhD. Thesis, Delft University of Technology, Delft University Press, 2004.
- [3] Cheung, C.Y. & Lam, W.H.K., Pedestrian route choices between escalator and stairway in MTR stations. *Journal of Transportation Engineering*, **124(3)**, pp. 277-285, 1998.
- [4] Van de Reijt, A. *Route choice model for passengers in railway stations*, MSc. Thesis, Delft University of Technology, 2004. (In Dutch.)
- [5] Daamen, W., Bovy P.H.L. & Hoogendoorn, S.P., Passenger route choice concerning level changes in railway stations, *PrePrints (CD-ROM) 84th Annual Meeting Transportation Research Board*, pp. 1-18, 2005.



How is the business case used by stakeholders for making project decisions with PFI/PPP projects?

M. J. Gannon

Westminster Business School, University of Westminster, London, UK

Abstract

In the United Kingdom (UK) Metro and Light Rail projects (new schemes, extensions and enhancements) are traditionally promoted by public sector sponsors within an extensive network of public and private sector stakeholders. This structure of public and private stakeholders has emerged since the structural and funding changes to the UK rail industry that occurred in the early 1990s; particularly the case with Public-Private-Partnership/Private Finance Initiative (PPP/PFI) rail projects. This complex stakeholder network includes: private sector infrastructure providers, train-operating companies; and public sector County Councils, Government departments and advisors and consultants from the private sector acting on their behalf.

Each group of project sponsors needs to address a range of complex project decisions throughout the project's development period. The business case provides decision makers with a tool to analyse and assess options to inform and justify the impact of these project decisions. Therefore the information contained within the business case is critical to facilitate successful project decision-making. Although literature provides a best practice for the business case form and an outline of expected project decisions, it provides minimal information regarding the nature of these decisions and how the business case and supporting information is used in addressing these decisions.

Keywords: PFI/PPP projects, business case, decision making, rail projects.



1 Introduction

Rail schemes in the United Kingdom (UK) often take longer to progress through the development cycle than project sponsors and the general public expected. Many schemes appear to remain on the “drawing board” in some cases for nearly a decade. Two fine examples of this are Croyley Rail Link [1] and Crossrail [2].

During the development stage, projects waiting for a “green light” to commence the planning process for a Transport and Works Act order (TWA) undergo a business case scrutiny check firstly by corporate decisions makers and then Government decision makers. During this second scrutiny by Transport authorities, projects undergo a barrage of business case related questions and enter into a seemingly perpetual cycle of questions and requests for project information from project sponsors and their advisors by the policy and decisions makers. The reason for this state of perpetual “development suspension” has two aspects. From the critics perspective it is due to the indecision of the policy and decision makers who sometimes appear to ask continual questions regarding the business case with perhaps a hope of delaying the approval decision; and from the policy and decision makers perspective it is due to the sponsor not providing sufficient information or in the form requested or modelling inconsistencies have been found and need resolving. Either way the project is on hold and becomes delayed in its fight for approval. In extreme cases structural changes have occurred to the industry during this time.

Similarly during the transaction stage projects can also be subject to major delays often bordering on having the project halted: due to political disagreement as was the case with London Underground’s Public-Private-Partnership (PPP); or project’s bid costs exceeding the Government’s expectations as was the case with three Light Rail Transit (LRT) UK schemes [3]: Leeds SuperTram (41% higher than expected), Manchester Metro (73% higher than expected) and South Hampshire Rapid Transit (59% higher than expected).

This prolonged wait for rail investment in the UK is sometimes not helped by the seemingly lack of decision making and unsuccessful decision making that occurs based on the information contained in a scheme’s business case [3]. Additionally when a decision is made, “go or no-go”, by the policy and decision makers, there is often limited transparency provided to project sponsors or the general public regarding how the decision was made. This is not helped when critical information on which the project decision was made is excluded from the business case framework.

The aim of the paper is to investigate using a qualitative Case Study and Survey research method how the business case and supporting information is used by public sector project stakeholders (decision makers) to support project decision making for PFI/PPP projects during the development phase. The investigation will focus on identifying key project decisions, the business case form and additional information used in the decision making process; and contrast these findings with the best practice business case form appropriate for decision making advocated in literature to identify any shortcomings with the business case that could be used to facilitate improved decision making with projects.



2 Public Finance Initiative/ Public-Private-Partnerships (PFI/PPP)

2.1 What is PFI/PPP?

According to Li et al [4] there are eight models of Public-Private-Partnerships (PPP's) of which PFI is the most commonly used form in the UK. PFI is essentially where the public sector contracts with a private sector consortium on a long-term basis often between 20 to 30 years to deliver services. During this period the private sector typically Designs, Builds, Finances, Operates and Maintains (DBFOM) the assets and provides a service to the client according to a performance specification. There is a close proximity between the PFI model and the PPP model as used to fund London's Underground system [5]; PPP in this case was used generically however was best described as an acquisition and a variation of the PFI model incorporating lessons learnt from the UK Rail privatisation.

The initiative has offered a means of funding large-scale public sector projects whilst not impacting on the Government's public sector borrowing requirement (PSBR) and has allowed private sector skills and innovations to be brought into the public sector to improve efficiency and provide Value for money (VfM). For contractors the initiative provides a 'golden opportunity' to secure long-term contracts with a steady payment stream in return for a delivered service.

2.2 Extent of PFI/PPP projects in the United Kingdom

Public-Private-Partnerships (PPP) bring together the public and private sectors in a long-term partnership. It has been an innovative means by which the public sector has attracted the private sector to invest in public services. The years 1987 to 2004 has witnessed nearly £40 billion of private capital invested in 626 projects across 20 public sectors departments, see Figure 1. Although prior to 1992 private investment in public sector projects was governed by the Ryrie Rules. The most significant of this expenditure and most complex projects have been within the transport sector where nearly 50% of this expenditure has occurred. With the Health sector witnessing the largest number of PFI projects signed in the UK.

At present PFI/PPP represents an extremely important method to Her Majesty's Treasury (HMT) of financing public sector projects. PFI has provided a means of reaching those parts of the public sector where privatisation was not possible whether this was due to social policy, lack of financial viability or where the full cost of the industry could not be met from the users or all of these [6]. Such examples where this initiative has been applied to investment in public services include: roads, hospitals, schools, prisons, defence, light rail and metro systems see Figure 1.



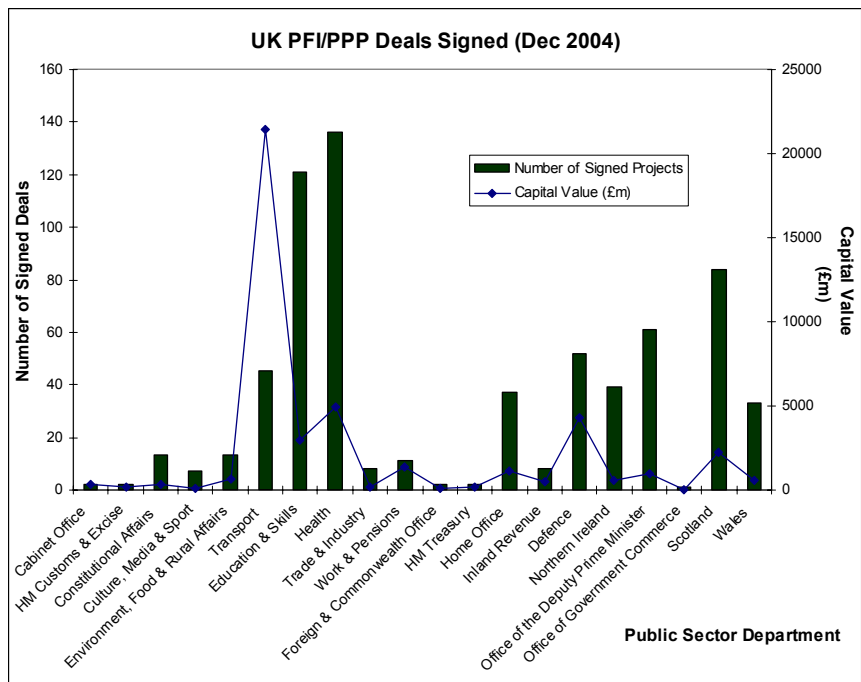


Figure 1: Number of PFI/PPP deals signed by public sector department in the UK Between 1987-2004 [7].

2.3 Project stakeholders in PFI/PPP projects

Project stakeholders refer to those groups who are positively or negatively affected by the delivery of the project [8]. Figure 2 illustrates the groups of project stakeholders for a rail project positioned against their project decision making authority with the business case and business case information available to the stakeholder on which to base their decision. Project sponsors and corporate sponsors are those groups on a project who have authority for decision making with the business case. Project sponsors usually make decisions during the development of the business case and corporate sponsors make strategic project decisions using the business case.

The project sponsors supported by their advisors are the best informed with business case information in terms of the development of the case and why decisions were taken at the project level. Whereas at the corporate level this “development” information is summarised and included within the business case for decision making purposes although not all of this information will be known on which the decision is made.

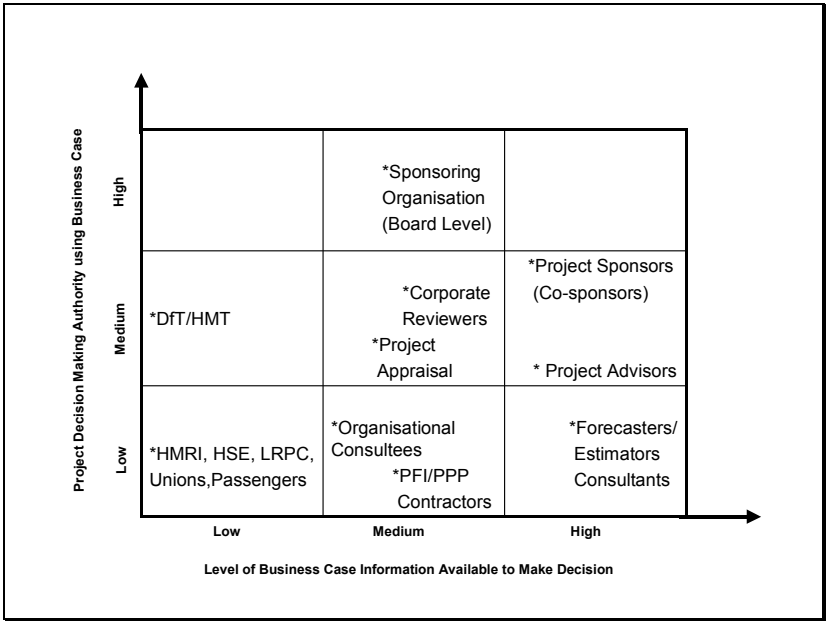


Figure 2: Stakeholders used in project decision making and availability of information on which to base decision.

3 The business case

3.1 Purpose

The purpose of the business case is well established by decision makers. A business case is used as a means of obtaining management commitment and approval for an investment providing business change through a clearly presented rational; and a framework (structure and contents) for informed decision making in planning and management of the business change and subsequent benefit realisation [9]. According to references [10] and [11] the purpose of the business case is to provide all the information needed to make an informed decision as to whether a project should be funded. However the business case in many organisations is limited to the quantitative aspects of the project rather than the wider project aspects on which the project decision is based.

3.2 A best practice framework

The OGC [9, 11] have suggested a general framework for the business case that comprises of five cases: linking the organisation and its corporate strategy to the project’s economics and finance and the organisation’s ability to deliver the project, see Table 1. The five cases include [12, 13]: The Strategic Case (Strategic fit), The Economic Case (Options appraisal), The Value for Money Case (Commercial aspects), The Financial Case (Affordability) and The Project

Management Case (Achievability). The Strategic case is equivalent to the Project Brief in PRINCE2TM and is often identified prior to the development of the project; the Economic Case and Project Management Case are part of the Project Information Document (PID) in PRINCE2TM.

Each of the five cases has elements that are essential to capture project information on which to make investment decisions [13]. Information relating to the corporate, finance and planning, aspects of a project is covered by Strategic Fit and Affordability. Information relating to the project itself is covered within the Options Appraisal, Commercial and Achievability. Commercial aspects are considered in the project's business case when the project is procured using PFI/PPP. Each of these cases seeks to establish whether the project: will meet the need of the business (Strategic Fit); progress the most appropriate option (Options appraisal); demonstrate the preferred option is fully achievable (Achievability) and the organisation fully understands the project implications; is affordable over the project's whole life especially including service changes (Affordability); and has sound commercial arrangements for external procurement and demonstrates value for money (Commercial). This framework can be used as a business case for all types and sizes of projects.

4 Research methodology

A qualitative case study and survey research strategy utilising PFI/PPP metro and light rail projects was used to investigate: project decisions made; information that was used to support these decisions; and information boundaries of the business case. The case studies were selected due to their position within the project development stage and also due to their differences in nature. For instance underground, surface line projects and light rail.

Data collection was by way of semi-structured interviews, transcribed, with project stakeholders; corporate finance, project sponsors and consultants who were responsible for developing cost and revenue forecasts for these projects to provide a context to the three areas under investigation. Project reports were reviewed to business case related information used in decision-making. This provided material from which key information themes were identified and analysed using within case analysis. The affordability decision was the main project decision investigated with each of the case studies.

5 PFI/PPP project decision making and information used

5.1 Project decision and business case form

The form of the business case specified in each of the cases investigated was dependent upon whether the investment, procurement or financing project decision was being addressed. Table 1 shows the 'Project Decision', the 'Business Case' form and the 'Purpose' of the business case [14]. With PFI/PPP projects the procurement and financing decision are the same decision. Figure 3 show the business case equations used to calculate each business case.

Each of the three business cases, shown in Table 1, use common cost and revenue components, with the exception of the economic case that uses social



benefits. In either case components are combined and it is their purpose that really distinguishes the form of each business case.

Table 1: Business cases required for PFI/PPP projects.

Project Decision	Business Case	Purpose
Is the investment Justified?	Economic case: Options Appraisal	To assess whether the Investment is worthwhile.
Does the investment provide value for money to the public sector implementing the project using PFI/PPP?	Value for Money (VfM): Commercial Aspects	To assess whether the project should be funded and procured by using PFI/PPP or traditionally (i.e. grant).
Is the project delivered using PFI/PPP affordability to the public sector?	Financial Case: Affordability	

(i) Economic Case: The economic case provided comfort for the economic justification of the project. This is required before proceeding with the PFI/PPP procurement route. It determines the benefit to society of carrying out the project. A key point with this analysis was that only incremental costs, revenues and benefits associated with the investment were considered. A benefit to cost ratio (BCR) was determined and used to assess whether the projects were a worthwhile investment.

(ii) Value for Money Case: Two business cases are prepared to determine Value for Money (VfM), the Private Finance Initiative (PFI) case and the Public Sector Comparator (PSC). Each of these cases were prepared to assess and compare the cost to the public sector, on a whole life basis, of undertaking the project under PFI/PPP or through traditional means using a 'hypothetical' comparator namely the PSC.

At the development stage the PFI business case was governed by the project's intended contractual model. This was comprised of the project's revenue, service charge costs (for a given level of service) and risks retained by the public sector. The service charge represents the private sectors revenue stream from the project. This was calculated by deriving the rate of return for the project from the project's cash flows; where the cash flows covered the private sector's capital and operating costs and risks for undertaking the projects. The PSC was a theoretical business case to determine the cost of carrying out the project using traditional procurement. This case assumed grant was not constrained and reflected the cost and the total risk to the public sector of implementing the project's performance specification.

(iii) Financial Case: This was used to assess the affordability of the project to the public sector and was an analysis of the case for incremental project revenue less incremental service charge incurred for the PFI/PPP project. The emphasis



was on determining the impact of the project on the gross margin and whether the project was affordable to sponsors. A project might well represent VFM however is unaffordable to the organisation. The project's cashflow statement was used to inform and justify the Net Financial Effect to the sponsoring organizations.

Economic Justification:

$$BCR = NPV[\text{Social Benefits}]/NPV[NFE]$$

If $BCR > \text{Threshold}$ investment justified.

Value for Money Decision:

$$VFM = NPV[\text{PSC Costs} + \text{Risk}] - NPV[\text{PFI Service Charges} + \text{Risk}]$$

If $VFM > 0$ Value for money has been achieved.

Affordability Decision:

$$NFE = NPV[\text{Revenues}] - NPV[\text{PFI Service Charges}]$$

Figure 3: PFI/PPP business case equations.

5.2 Information used in decision making

Although not considered as part of the business case by the sponsors wider project information was also considered as part of the affordability decision making process. A sample of this information used to support this decision making is shown in Table 2. With the cases examined the business case was often synonymous with an appraisal and hard information (quantitative) rather than soft information (qualitative). The business case information was represented at two levels the Corporate decision making level and the Project level; containing the financial model and the detailed development knowledge of the project and business case.

6 Shortcomings of the business case

The project sponsor's business case predominantly contained hard financial information therefore not surprisingly its role was restricted to the financial analysis that played a minimal role in project decision making. It was evident significant soft information was utilised in the project decision making process outside the sponsor's business case, see Table 2. This exclusion of information from the business case removed transparency in the decision making process.

Equally the business case used for decision making at the corporate level provided no indication to decision makers of the development process. For instance nothing is brought to the corporate decision makers attention in a



business case framework regarding model calibration or validation, forecasting methodology, capability and experience of forecasters or data sources. These elements are often researched by the project sponsor before making a corporate decision but not included within the business case framework to facilitate good practice.

Table 2: Hard and soft information used by sponsors in project decision making.

Information Used in Project Decision Making	
Hard Information	Cashflow and Net financial effect on sponsor (annual and PV) for central case.
	Cashflow with scenarios
	Cashflow with sensitivities
	Benchmarking outturn capital and operating costs with project costs.
	Benchmarking operating costs and revenues with operating system.
	Benchmark revenue forecasts against revenue forecasts from other models.
	Quantitative risk register
Soft Information	Sponsor's statutory obligations
	Sponsors' existing project commitments and priorities
	Co-sponsors business need and urgency
	Co-sponsors financial return
	Co-sponsors relationship with consultants and information control
	Government Authorities project expectation and position
	Political support for project (MP's, Mayor)
	Opinions of senior stakeholders
	Sponsor's long term policies (fares and operations)
	Five year budgeted amount for project
	Existing board papers and agreed status of project
	Lessons learnt on previous projects
	Judgement of bias of information
	Discussions/ agreements made at Senior Management level
	Contractual model precedents
	Contractual principles (proposed or existing)
	Qualitative risk register

Having investigated the case studies there is an argument to widen the boundaries of the business case to include this information within the case to provide transparency with decision-making. Additionally there is an argument to deepen the business case to capture critical business case development information from the project team for corporate level decision makers thus providing a framework to facilitate improved decision making.

7 Conclusion

Best practice business case frameworks provide a useful template for the business case however there are information shortcomings in terms of its width and depth. Investigating the case studies has demonstrated the business case is often associated with an appraisal and its role limited to quantitative aspects supported by wider project information used in decision-making. However this wider information used for decision-making is not cited as part of the business case framework although is critical information.



To provide transparency in decision making it is essential all information used in the process is included within the business case framework. By providing an additional detailed framework in which to capture business case development information will provide decision makers with further assurance decisions are made on a credible business case. Supplementing this form of business case with lessons learnt from past projects will provide a business case framework that would facilitate improved decision making with projects thereby reducing wasted costs and time. The development of a new business case framework to facilitate improved project decision-making is currently being researched at the University of Leeds by the author.

The contents of this paper represent the views of the author and do not represent University of Westminster's policy. The author does not accept any liability for its correctness

References

- [1] Beherman, D. (2004). Rail link moves a step closer to being realised. Watford Observer. Watford: p3.
- [2] Murray, D. and R. Lydall (2004). After 15 years no Crossrail - but an astonishing £30m-a-year bill. Evening Standard. London: p8-9.
- [3] HC (2005) House of Commons, Transport Committee Integrated Transport: The Future of Light Rail and Modern Trams in the United Kingdom, HC 378-1, 3 April 2005.
- [4] Li Bing, A. Akingtoye, P.J. Edwards and C Hardcastle (2005), Critical Success Factors for PFI/PPP Projects in the UK Construction Industry, Construction Management and Economics (June 2005), 23, p459-471.
- [5] Gannon (2002). Funding London's Underground Using PPP? European Transport Conference, Homerton College, Cambridge University.
- [6] Morton, A. S. (1998). "The Developing Initiative." Private Finance initiative Journal 3(2): p6-7.
- [7] HMT (2004). PFI/PPP Signed deals up to 2004.
- [8] Gardiner Paul. D Project Management: A Strategic Planning Approach, Pelgrave Macmillian, 2005.
- [9] OGC (2002) Office for Government and Commerce, Successful Delivery Toolkit, www.ogc.gov.uk/sdtoolkit/bclifecycle/b/b2.html.
- [10] Schmidt, M.J 1999 What's a Business Case? And Other Frequently Asked Questions. Solution Matrix, Ltd: www.solutionmatrix.com.
- [11] Marchewka, J.T Information Technology Project Management, Second Edition, John Wiley and Sons, 2006.
- [12] OGC (2003) Office for Government and Commerce, Business Case Template, www.ogc.gov.uk/sdtoolkit/reference/documentation/p13-buscase.html (March).
- [13] OGC (2004) Office for Government and Commerce, Business Case Development Overview, www.ogc.gov.uk/sdtkdev/examples/dwp_life/prodsets/strands/fbc_hgc.doc.
- [14] Gannon (2004) A Review of Techniques Used to Assess the Credibility of a Business Case in Public Sector Rail Projects, COMPRAIL2004, Dresden.



A bilevel model for optimizing station locations along a rail transit line

S. Samanta & M. K. Jha

*Department of Civil Engineering, Morgan State University,
Maryland, USA*

Abstract

The population growth and increase in number of commuters in urban areas give rise to the need to either build or extend a rail transit system requiring: (i) optimization of the alignment of a rail transit line and (ii) optimization of station locations along rail transit lines. The positions of the stations depend on many factors, such as total cost for locating them, proximity from the residential neighborhoods, feasibility studies, and environmental, and political factors. In this paper a bilevel programming approach, which minimizes the total cost, is proposed to solve the station location problem. At the lower level, the potential ridership generated from the major cities or CBDs are estimated by dividing the study area in optimum number of zones, to maximize the usage by the potential riders. At the upper level, the number and location of intermediate stations are determined by minimizing the sum of user, operator, and location costs. The primary components of costs are the costs associated with traveling to the station and in-vehicle travel time, land-acquisition cost (also known as right-of-way cost), and cost of operating the train, construction of stations and parking facilities. The total cost is minimized using a Genetic Algorithm (GA). A Geographic Information System (GIS) is used in order to work directly with maps of the proposed rail-line, existing road networks and transit lines, and land and property boundaries. The population and passenger distribution in the study area as well as the travel times are obtained using a GIS which is integrated to the GA, to obtain the best station locations. The model is applied to decide on the optimum positions of stations along a transit corridor of an artificial case study. The results indicate that one can optimally locate station locations with improved precision if GIS data with sufficient accuracy were available.

Keywords: bilevel programming, rail transit station location, genetic algorithm.



1 Introduction

The population growth and the increase in the number of commuters in urban areas give rise to the need of extending the facilities of the existing transportation system. Transit system has been considered to be an important and efficient transportation mode for the mobility of mass population. It is gaining preference these days as it discourages the congestion and pollution, generally caused by automobiles. A rail transit system comprises of a transit line and stations along it. Once the transit line is aligned, the issue of locating the stations functionally and optimally along it arises. The positions of the stations depend on a number of factors, including the total cost for locating them, a feasibility study, and environmental and political factors. But, the most significant factor which influences the location of a station is the travel demand at the particular zone surrounding the station. In order to estimate the travel demand for a transit system potential rail transit riders have to be identified. The transit ridership or the travel demand generated at a particular station or zone is governed by the socioeconomic and demographic factors of that geographic area. The characteristics of the regional population play important roles while estimating the travel demand.

The research problem in the paper is to find the optimal number and locations of stations along a rail transit line provided the source, destination and the alignment of the transit line are given. The model is developed based on the potential ridership along the transit line. The objective function of a station location problem is developed as a cost minimization problem, which is a non-linear, and non-convex programming problem. The problem is formulated as a bilevel programming model. The concept of bilevel programming is applicable when two different stages of the problem share the common set of resources, but control different sets of decision variables to reach to the final solution. The station location problem is modeled such that the clustering of population is done homogeneously at the lower level, in order to obtain the optimal services from the stations, and at the upper level the cost is minimized, to optimize the number and location of stations. The population data can be obtained using GIS database. A Genetic Algorithm (GA) is used to solve the optimization problem at the upper level. The model is applied on a small artificial case study, where the optimal locations of stations are obtained after the total population along the transit line is clustered in two groups.

2 Literature review

The station location problem along the transit line has been approached by various researchers in different times. The problem of locating stations along transit rail lines in urban areas seems to be interesting and challenging to the researchers from the earlier days. Vuchic and Newell [1] studied the station location problem considering the linear transit line and uniform passenger distribution along it. They solved a set of partial differential equations to minimize the total cost. Wirasinghe [2] considered rail-station location as one of



the parameters to design nearly optimal parameters for a rail/feeder-bus system on a rectangular grid. The approach to solve the problem was to develop an analytical model to minimize the travel time and the operating costs by considering a rectangular grid, which is not applicable in practical life. Wirasinghe and Seneviratne [3] tried to optimize rail line length in an urban transportation corridor by minimizing total transportation costs. Formulation of the various cost components, e.g., user and operator costs were described vividly in their study. Spasovic and Schonfeld [4] developed an optimization model for transit service coverage using a penalty method. The stop spacing was obtained by considering a differentiable cost function. Kikuchi and Vuchic [5] developed a process which provides a theoretical basis for the selection for transit stopping policy and the number of stops. This study does not specify the spacings between them or the locations along a transit line. Horner and Grubestic [6] proposed a planning approach for locating rail terminals based on GIS data. It is mainly based on the demand parameter rather than the cost consideration. In another study by Lutin and Markowicz [7] the potential ridership was estimated considering the corridors based on the residential landuses. The demand was considered to be uniform. The cluster was developed based on the density and the proximity of the transit station in that study.

Bilevel Programming concept, which is fairly new, is applied on various transportation problems. Le et al. [8] formulated a bilevel model for the terminal location for a logistics problem. They designed the lower level model as per the behavior of the individual vehicle using the terminal minimizing its own cost and the upper level as the behavior of the planner for minimizing the total cost of the logistics system and the optimal size and location of the terminals. A GA is used to solve the problem optimally. Huang and Liu [9] proposed a bilevel approach for a logistics distribution network with balancing requirements. It promises to give a flexible network. The upper level is to minimize the total cost and the lower level is designed to obtain the balanced workload for the network. A GA is used to solve the problem. A bilevel program model is designed for toll optimization on a multicommodity transportation network [10]. At the upper level, the objective is set to maximize the revenue and the lower level is to minimize the total travel cost, i.e., to obtain the shortest path in the network. The model is applied on two examples. Murray-Tuite and Mahmassani [11] developed a bilevel formulation to determine the vulnerable links in a transportation network. At the lower level, the traffic assignment is done optimally and the vehicles are assigned accordingly. At the upper level, the disruption is maximized in the network based on the vulnerability index. The classical Traveling Salesman Problem (TSP) is also solved by bilevel programming [12]. The toll optimization problem is reduced to a TSP problem and is solved by relaxing the bounds. Hejazi et al. [13] and Yin [14] proposed a GA based solution tool for bilevel programming models. Mathur and Puri [15] studied the bilevel programming model for a bottleneck of non-convex problem. They studied the nature of the problem and proposed a formulation and solution methodology. By reviewing the literature the motivation to apply the bilevel



programming on the station location problem is drawn in order to incorporate the potential ridership in the model.

3 Model formulation

The bilevel programming model is designed as a two-stage optimization problem, where the optimal values of one set of variables obtained at one level are used to determine another set of variables at another level. The representation of a typical bilevel programming can be found in the literature [9]. The station location problem is formulated as a bilevel programming model as given below.

The formulation of the objective function and constraints are shown in two different levels, such as lower and upper levels. The description of the formulation of the model is represented by fig. 1. The structure of the bilevel programming model is shown in fig. 2.

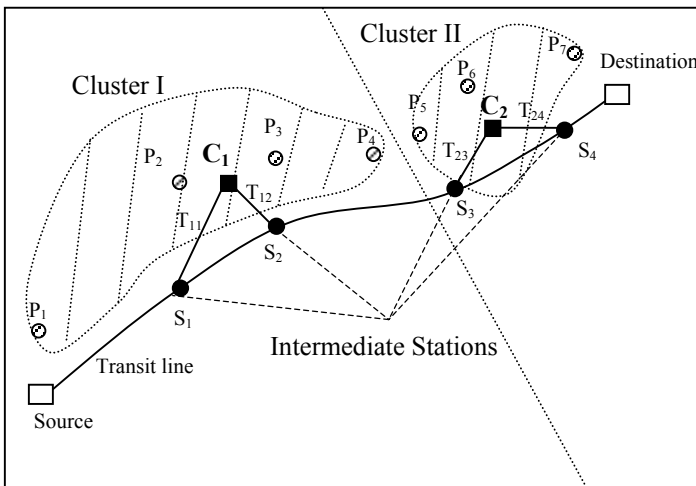


Figure 1: An example showing the locations of intermediate stations, clustering of population and the centroids.

3.1 Lower level

At this level the aim is to cluster the population of the study area into a number of homogeneous subgroups using the concept of classification [16]. The centroids are obtained by clustering. The equivalent clustered population of each subgroup is supposed to act from the respective centroid. The travel time of the clustered population is calculated from the centroid of each subgroup to the stations which remain within the range of the particular cluster. The study area and the length of the rail transit line are divided into the equal sub areas or segments, respectively as that of the number of subgroups. The mapping is done to the centroid of a subgroup and one or more stations, which remain within the



respective segment of the transit line, respectively and the travel time is calculated from the centroid to the respective stations. The total population is partitioned in K exclusive, and collectively exhaustive subgroups. The homogeneity of the subgroups will depend on the variance of their population values. After creating the subgroups the equivalent population of the subgroup and the positions of the centroids of the subgroups are calculated by the centre-of-gravity method.

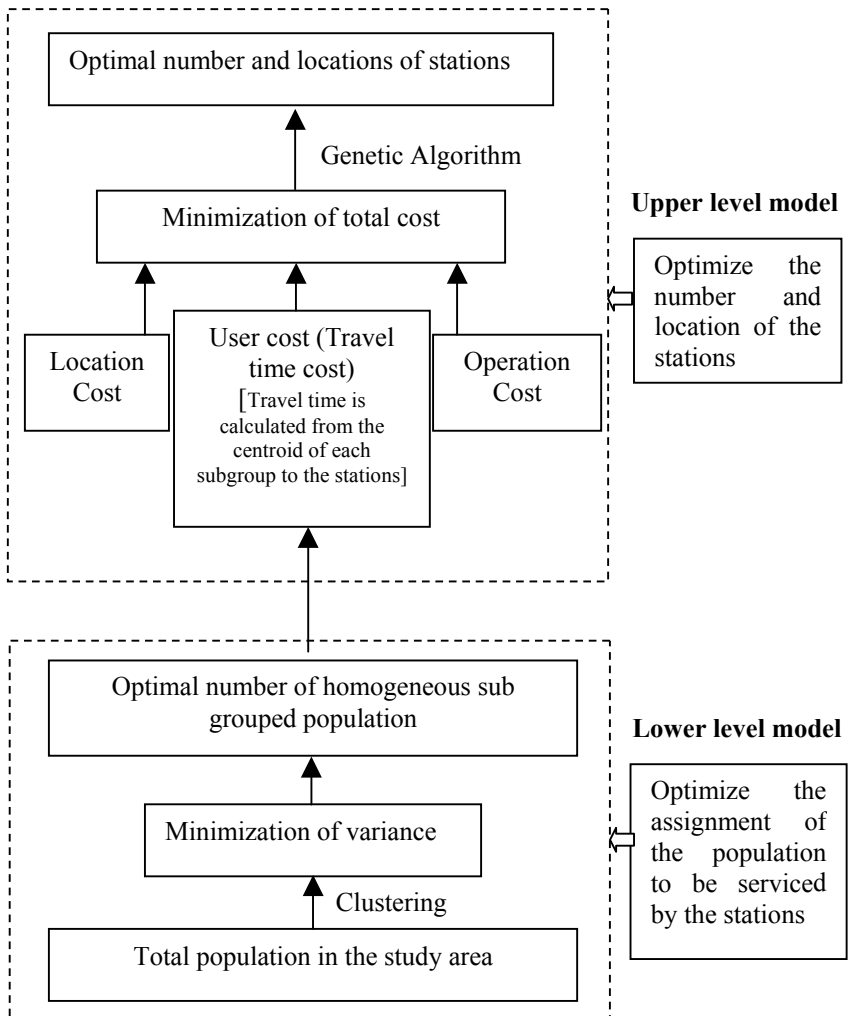


Figure 2: Structure of the bilevel programming model.

Each group is represented by:

$$X_k : (G_{C_k}, D_{C_k}); k = 1, 2, \dots, K$$



where, G_{C_k} = equivalent population of k^{th} subgroup

D_{C_k} = position of the centroid of k^{th} subgroup

The objective function is expressed as

$$\text{Minimize } V_K \quad (1)$$

Subject to

$$K \leq N ; \text{ if } D_i - D_{i-1} \geq \frac{L}{N} \quad (2)$$

$$K \leq \frac{L}{N - J} ; \text{ otherwise,} \quad (3)$$

where, J = number of times when $D_i - D_{i-1} \leq \frac{L}{N}$

where, V_K is the variance of the population values of the K subgroups given by,

$$V_K = \text{Var}(\overline{G_C})$$

where, $\overline{G_C} : \{G_{C_k}\}; k = 1, 2, \dots, K$

3.2 Upper level

At upper level, the total cost is minimized based on the division of the study area in optimal number of subgroups and the travel time cost of the population of the study area. The objective function can be written as:

$$\text{Minimize } TC = UC + OC + LC \quad (4)$$

Subject to

$$\Delta S_i \geq \Delta S_{\min} \quad (5)$$

where, UC = User cost = unit travel cost (UTC) $\times$ total travel time

OC = operator cost = unit operator cost (UOC) $\times$ vehicle travel time (t_v)

LC = Location cost (property value)

ΔS_i = Spacing between i^{th} and $(i+1)^{th}$ stations

$\Delta S_{\min}$ = Minimum spacing

The above costs have to be formulated as a function of the decision variables which are the coordinates of the station points. Mathematically, the cost can be formulated as,

$$TC(D_1, D_2, \dots, D_N) = \sum_{k=1}^N TC(D_k) = \sum_{k=1}^N \left[\sum_{i=1}^K UC(D_k, P_i) + OC(D_k) + LC(D_k) \right] \quad (6)$$

where, $D_k = k^{th}$ station

P_i = centroid of i^{th} subgroup



Now the passenger access time (t_a) is calculated as it plays the most important role in the total cost calculation. For i^{th} subgroups and N^{th} station, the total access time of the passengers can be written as,

$$t_{a_i} = \sum_{j=1}^M p_i \times t_{ij}; i = 1, 2, 3, \dots, K \quad (7)$$

where, p_i = population of i^{th} cluster

t_{ij} = travel time from the centroid P_i to j^{th} station

M = the number of stations which exist within the range of the i^{th} cluster

The operator cost is a function of the distance traveled by the vehicle, and the location cost is composed of the cost of right-of-way acquisition.

4 Solution methodology

The optimization problem is solved using a GA based algorithm, which is modified from the previous work by Jha and Oluokun [17].

4.1 Steps of the genetic algorithm

Step I. Generate a random number between 1 and N : $r_d(1, N) = k$; k is a positive integer. This is the number of intermediate stations between the source and destination.

Step II. Generate k random numbers $(r_{d1}, r_{d2}, \dots, r_{dk}) = [\lambda_1, \lambda_2, \dots, \lambda_k]$ in the interval $(\Delta S_{\min}, L - \Delta S_{\min})$, which represent the interstation spacings. The initial population consists of $\lambda_1, \lambda_2, \dots, \lambda_n$ which represent the distances of the possible intermediate stations from the starting station.

Step III. Calculate fitness (total cost function) of the population members.

Step IV. Apply mutation and crossover operators. Mutation operator is developed by randomly selecting a gene and replacing it by a randomly selected real number within the limiting values. The mutated chromosome becomes $[\lambda_1, \lambda_2, \dots, \lambda'_k, \dots, \lambda_n]$, where λ'_k is the replaced value (changed distance from the starting station).

Step V. Develop the crossover operator by selecting two solutions from the population randomly. A part containing one or more than one station is identified from each of these two solutions and they are exchanged maintaining the feasibility conditions to produce two new solutions.

Step VI. Develop a selection-replacement scheme to ensure efficient convergence towards the global optimum.

Step VII. Iterate for the specified number of generations from steps I-VI.

Step VIII. Stop when searching through specified number of generations is complete or improvement in the objective function value is negligible (within 1%). Obtain the optimal station sequence and associated optimal cost.



Table 1: Population of cities.

City	Population
P_1	10,000
P_2	5,000
P_3	25,000
P_4	10,000
P_5	18,000
P_6	16,000
P_7	13,000

Table 2: Location cost for the stations along the transit line.

Distance from the source (ft)	Cost (\$)
0	10,000
2	10,000
4	25,000
6	17,200
8	13,000
10	15,600
12	12,000
14	30,000
16	22,000
18	15,300
20	14,500
22	16,900
24	15,000

5 Numerical example

The model is applied in a small artificial case study. It is assumed that seven cities e.g., $P_1, P_2, \dots, P_7$ are present in the study area (Figure 3). For simplicity, the transit line is assumed to be linear and the coordinates of the source and destination are given. Based on the population (Table 1) and the location of the cities we cluster the total population in two mutually exhaustive subgroups. The clusters are obtained by minimizing the variance of the total population of the two subgroups at the lower level. We find out the centroids of these two subgroups, which are represented by C_1 consisting of the clusters of P_1, P_2, P_3 and P_4 and C_2 consisting of the clusters of P_5, P_6 and P_7 . The populations of the two clusters are 50,000 and 47,000 respectively. The coordinates of C_1 and C_2 are given by (10.4, 18.5) and (18.9, 21) respectively. The total population of each cluster is assumed to act from the centroid and the positions of the centroids are shown in Figure 3. Thus, we obtain two homogeneous subgroups of population from which the travel time cost is calculated to the stations along the transit line. The stations are allocated to the subgroups for the service based on their locations and the range of the cluster under which they fall. The unit travel time cost and unit operator cost considered for the example are 1000 dollars/ft



and 2000 dollars/ft respectively. The location costs are given in Table 2. We calculate the travel time to the stations from the centroid of the cluster within each subgroup. We add the operation cost and location cost along with the calculated travel time cost to constitute the total cost. At the upper level we minimize the total cost using the Genetic Algorithm. We obtain the location of four stations S_1 , S_2 , S_3 and S_4 (Figure 3) which gives the minimum cost of $\$57.603 \times 10^5$.

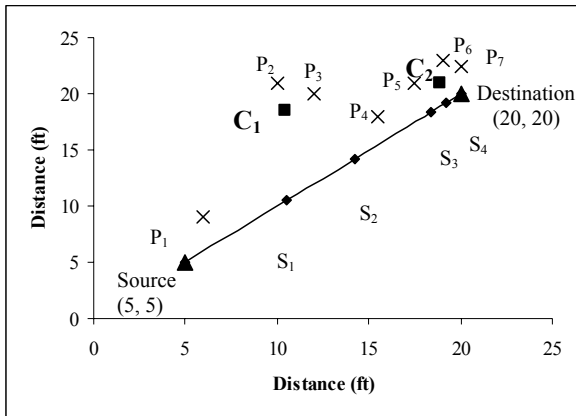


Figure 3: Optimal locations of stations.

6 Conclusion and future scope

The bilevel programming model is applied on a small artificial case to determine the optimal number and locations of stations along a rail transit line. The clustering of the population is done at the lower level. The efficiency of the model depends on the accuracy of the clustering process. The intent of clustering is to distribute the population homogeneously to the stations for obtaining the optimal service from the stations. For the example studied, we obtain two clusters of population at the lower level and four stations along the transit line at the upper level as optimal solutions. It has been observed that as the clustering varies, it gives different solutions. As we increase the number of clusters, the variance decreases but the total cost increases. Additional research has to be carried out to explore other variables which can make the clustering process perform better and more efficiently, other than the variance. The parametric study has not been considered in the scope of the present study. The sensitivity analysis of the parameters has to be performed to make the model robust. The model has to be validated on a real life case as in future works.

References

- [1] Vuchic, V.R. & Newell, G.F., Rapid Transit Interstation Spacing for Minimum Travel Time. *Transportation Science*, **2(4)**, pp. 359-374, 1968.



- [2] Wirasinghe, S.C., Nearly Optimal Parameters for a Rail/Feeder-Bus System on a Rectangular Grid. *Transportation Research-A*, **14**, pp. 33-40, 1980.
- [3] Wirasinghe, S.C. & Seneviratne, P., Rail Line Length in an Urban Transportation Corridor. *Transportation Science*, **20(4)**, pp. 237-245, 1986.
- [4] Spasovic, L. & Schonfeld, P., Method for Optimizing Transit Service Coverage. *Transportation Research Record*, 1402, pp. 28-39, 1997.
- [5] Kikuchi, S. & Vuchic, V.R., Transit Vehicle Stopping Regimes and Spacings. *Transportation Science*, **16(3)**, pp. 311-331, 1982.
- [6] Horner, M. W. & Grubescic T.W., A GIS-based planning approach to locating urban rail terminals. *Transportation*, **28**, pp. 55-77, 2001.
- [7] Lutin, J. M., & Markowicz, B. P., Estimating the effects of residential joint-development policies on rail transit ridership. *Transportation Research Record*, 908, pp. 7-12, 1983.
- [8] Le, Y., L. Miao, Wang H. & Wang C., A Bilevel Programming model and a solution method for public logistics terminal planning. Transportation Research Board, Washington DC, 2006.
- [9] Huang B. & Liu N., Bilevel Programming Approach to Optimizing a Logistic Distribution Network with Balancing Requirements. *Transportation Research Record*, 1894, pp. 188-197, 2004.
- [10] Brtcorne L., Labbe M., Marcotte P. & Savard G., A Bilevel Model for Toll Optimization on a Multicommodity Transportation Network. *Transportation Science*, **35(4)**, pp. 345-358, 2001.
- [11] Murray-Tuite P.M. & Mahmassani H.S., Methodology for Determining Vulnerable Links in a Transportation Network. *Transportation Research Record*, 1882, pp. 88-96, 2004.
- [12] Marcotte P., Savard G. & Semet F., A Bilevel Programming Approach to the Travelling Salesman Problem. *Operation Research Letters*, **32**, pp. 240-248, 2004.
- [13] Hejazi S. R., Memariani A., Jahanshahloo G. & Sepehri M. M., Linear Bilevel Programming Solution by Genetic Algorithm, *Computers and Operation Research*, **29**, pp. 1913-1925, 2002.
- [14] Yin Y., Genetic-Algorithms-Based approach for Bilevel Programming Models. *Journal of Transportation Engineering*, pp. 115-120, 2000.
- [15] Mathur K. & Puri M.C., Theory and methodology: A Bilevel Bottleneck Programming Problem. *European Journal of Operation Research*, **85**, pp. 337-344, 2005.
- [16] Ben-Akiva & M., Lerman S. R., *Discrete Choice Analysis*, The MIT Press, Cambridge, Massachusetts, London, England, 1985.
- [17] Jha, M.K. & Oluokun C., Optimizing Station Locations along Transit Rail Lines with Geographic Information Systems and Artificial Intelligence, in *Computers in Railways IX (COMPRAIL 2004)*, J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto, S. Sone (eds.), WIT Press, Southampton, U.K., 2004.



Design and implementation of virtual environments for planning and building sustainable railway transit systems

M. Chandramouli¹, B. Huang¹, T.-Y. Chou², L. K. Chung²
& Q. Wu¹

¹*Department of Geomatics Engineering, University of Calgary, Canada*

²*GIS Research Center, Feng Chia University, Taiwan*

Abstract

This article explores the application of virtual reality techniques to rail transit systems from the design and training perspective. We present virtual environments for interactive 3D visualization of a rail transit station. We describe the design and implementation of some 3D models, which offer a good level of user-interaction and animation within a common rail transit system. The virtual environments for interactive 3D visualization are modeled using Virtual Reality Modeling Language (VRML), a web standard for creating virtual worlds.

Two basic categories of models are discussed in our study. The static 3D models, which are used for planning and design purposes, are based on the object-oriented approach. The dynamic models, which are used for simulation and driver-training purposes, are based on the event-driven approach. Even though a standalone VR application is not capable of a higher level of interaction or animation, using SCRIPT nodes, JavaScript can be embedded in the program to provide additional functionalities. JavaScripting can be used ingeniously to manipulate and control some of the key animation and interpolation nodes to provide powerful functionalities within the VR world. In this study, we integrate the script programming language with virtual reality and elucidate simulations that can be employed for driver training purposes. By employing these visualization and simulation techniques, designers, engineers, planners, and decision-makers can assess the plan beforehand. The virtual worlds serve as immensely useful tools during the design process by helping the designers to position themselves in photo-realistic 3D immersive environments wherefrom the actual scenario can be viewed and modifications made accordingly.

Keywords: virtual environments visualization, animation, rail transit systems, object-oriented approach, event-driven approach, object linking and embedding.



1 Introduction

These days, rail stations must cater to the demands of a considerably larger group of commuters with varying needs and demands. Hence, considerable care has to be taken in planning and designing the railway stations. Visualization is a very useful tool that holds immense potential in planning and designing modern rail transit stations. Visualization enables seeing beyond the individual objects or components of the scene and facilitates understanding the entire scene as an interlinked framework. In real world systems, objects do not exist in complete isolation, but are related to each other, either directly or indirectly. Visualization is a tool of immense value that helps to tackle design problems and facilitates finding an optimal solution, which is both time-saving and economical. Visualization enables viewing the constituent elements of a rail -infrastructure framework in whole as well as in parts. This kind of object-based visualization is of significant value to designers, planners and decision makers.

2 VRML and 3D Visualization

A wide range of visualization techniques are available as on date. In fact, any pictorial representation can be considered a visualization tool. Today, one term that is synonymous with visualization over the web is 'Virtual Reality'. Much of the 3D visualization for a wide range of applications is done in a virtual space that is often characterized as 'virtual worlds'. The reason they are called virtual worlds is that they are not actually 3D worlds in real space, but they are digital or cyber worlds that have their own coordinate systems and define a 3D virtual coordinate space in which applications can be built. The users can navigate within these virtual worlds, interact with the objects, rotate or scale them, and transform them in multiple ways. These virtual worlds facilitate user interaction with the 3D objects and provide a sense of immersion.

Modeling using virtual environments facilitates visualizing the finished product (constructed rail station) before hand. When the designers, planners, and other possible members who are involved in decision-making view such a model, it facilitates non-destructive testing [1]. If, after construction, the decision-makers feel that the location of terminal facilities is not appropriate or the positions of some architectural elements have to be changed for aesthetic purposes, this would involve a lot of wastage of time and resources. However, a 3D immersive virtual environment enables visualizing the station even before it is built and the planners and decision-makers can view and navigate through the model at the convenience of an office desktop and changes can be made without wasting any money. The model can be viewed, discussions held, modifications can be finalized and these can be incorporated before the actual construction can begin. One recent trend is to incorporate public-participation in infrastructure projects. In such cases, the model can be hosted online and members of the public can walk-through the model and suggest possible modifications. Suitable suggestions can be incorporated into the model before actual construction can begin.



3 Static 3D models: Rail Transit Systems

This section focuses on the application of VRML for planning and designing rail transit stations using static models. The basic difference between a static and a dynamic model is that the objects in a static model do not have animation capabilities. Using static 3D models, it is possible to test an application before implementation and virtual worlds can be used to systematically construct rail transit stations [2]. More importantly such environments allow the design of transit stations according to diverse and changeable real-world requirements. The salient advantages of such 3D visual representations are that

- Multiple scenarios can be evaluated
- Infinite viewpoints can be generated
- Decision-makers can view the finished product before hand
- The virtual models are extremely time-saving and economical
- Visibility of signals and various other elements can be tested

3.1 Planning and design of terminal stations using object oriented approach

Designing modern railway stations is a complex process and the following important factors must be considered and incorporated into the planning and design process [3]. Today's rail stations are more complicated owing to the following reasons:

- Rail stations handle considerably greater number of trains per day
- Passenger numbers have increased dramatically over the past several years
- Many rail terminals are supposed to cater to multi-modal transport demands
- Stakeholders in station construction includes both public and private parties
- The financial investments in station infrastructure have increased manifold

3.1.1 Major elements of a rail station

Some common components of stations, when viewed as a collection of objects, besides platforms and the tracks, are:

- Public Telephones
- Vending stalls and/or Machines
- Toilets
- Ticket halls
- Luggage handling/storage rooms
- Station management room
- Fire-fighting or emergency equipments
- Stairs, elevators, and escalators

This is by no means an exhaustive list, as the station components vary depending on the complexity of operations and the level of commuter service. Nevertheless, in today's scenario stations are in a position to include more and more advanced equipments and technologies that can offer convenient facilities to the commuters. In such a situation the challenge is to properly position the existing infrastructure facilities in an optimal manner to best serve the commuters.



3.1.2 Design criteria

Stations can be defined as the links where the rail transit system establishes contact with the neighbouring regions. Rail terminal stations are also the points where commuters get access to the services and change over to other modes of transport. Therefore, meticulous station design and planning for optimal use of the system is inevitable. The primary objectives in the design and planning of a station must inevitably pay utmost care to include

- Aesthetic look
- Provisions for emergency management
- Adequate quantity and quality of facilities

The mere provision of space required for the infrastructure cannot satisfy the aforementioned goals. More than the amount of space available, it is the ingenious utilization of this space that determines the efficiency and performance of a station. The consumers i.e. the commuters might have diverse observations spaces [4] in a platform and such views should be able to elicit the amenities available and hence provide a wholesome glimpse of the station. One important consideration while including various amenities in a station should be that the several shops and booths should be carefully located so as not to impede commuter movement, whilst at the same time not located on an island within the station so that commuters who need services urgently can not access them. Currently, station design and approval might involve experts and specialists from diverse backgrounds, such as designers, engineers, technicians, architects, and even economists. In order for the designers and architects to be able to explain their schema to all the members clearly, they need to make use of a tool such as visualization. 3D visual worlds are excellent tools for communication and can be easily understood by a broad spectrum of people.

3.2 Station design using VRML

VRML adopts a hierarchical structure for describing 3D worlds. Visualizing the world as a group of objects within a structural framework facilitates modeling real-world scenarios better [2]. Typically, the elements mentioned earlier under section 3.1.1 can be found in a rail station. The station can be considered the root object that is at the top of the hierarchy, under which all other objects fall. There can be further ramifications depending on the object's complexity. For instance, platform is a child of 'Rail Station' and 'Ticket Hall' is a child of 'Platform' object (Figure 1).

There are several advantages of such grouping of elements composing a scene in the form of objects that have group or parent-child relationships. First and foremost, the overall organization becomes lucid and basically no element, however large or small, is left out. This facilitates quantity estimation greatly and also aids locating system components within a huge framework.

Another prominent feature offered by the object-oriented approach is that of 'reusable software objects' (Figure 2). Quite frequently, a single element is used more than once within a parent or group object. For instance, wheels of locomotives, sleepers used in the track system, inventory elements such as



lamps, signposts, furniture elements etc. In such a case, the object that occurs repetitively can be defined just once and instances of the object can be used subsequently.

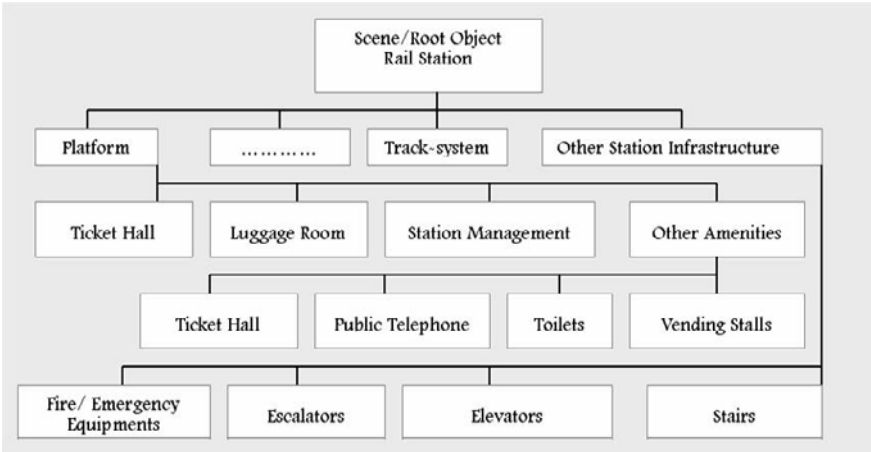


Figure 1: Hierarchical grouping of Station object.

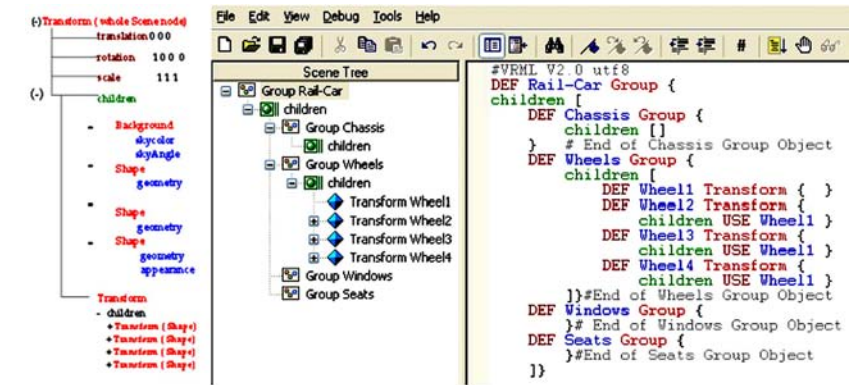


Figure 2: a. Scene Tree structure showing ‘Parent-Child’ relationships.
b. Hierarchical grouping of station elements.

Every component within a rail terminal has its unique function. For instance, the ultimate objective of signboard is to convey relevant and useful information to the commuters. This is lost if a signboard is improperly positioned or is in a condition wherein reading any information from it is difficult. Particularly, in today’s scenario wherein international tourists constitute a considerable chunk of commuters in metropolitan rail station, signboards are a key source of information. Virtual worlds serve as an excellent means to view the model of the station and its components from numerous perspectives, and optimally positioning the signboards. Similarly, during the planning stage, the various other elements can be placed in different positions at different locations and thus

can be seen from different perspectives and even from different heights. This will facilitate locating intruding objects or judging visibility.

3.3 User interface for multi-perspective and multi-scenario visualization

Figure 3 illustrates a VB powered GUI (Graphical User Interface) that facilitates multiple-scenario visualization with interactive functionalities. By means of OLE (Object linking and Embedding), the various virtual worlds are embedded and linked on a common interface. Each scenario depicts a different configuration of the same rail station, which can be a simple re-organization of the elements in the previous scenario or a total rearrangement of the station components.

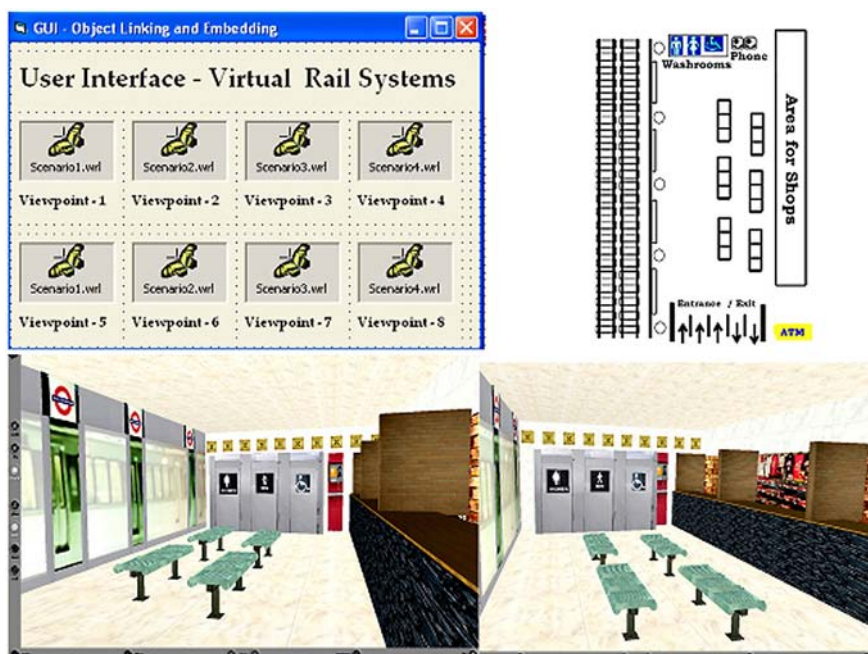


Figure 3: a. Visual basic GUI with Object linking and Embedding (OLE) functionality. b. Rough Layout. c. Viewpoint 1. d. Viewpoint 2.

Such virtual environments can also be used to create sophisticated rail car and locomotive models [5]. A very high level of detail can be built into the design of such 3D models and the dynamic functioning of the locomotive components can be visualized. Under real-world circumstances, it is practically impossible to get to the level of the individual components each and every time when it is required to visualize changes made. Using VRML highly detailed models capable of dynamic behaviour can be generated and more importantly these models are scalable. More importantly, the training can be incorporated even when the model is being designed. Some other key functionalities that can be added into the virtual worlds are described in Figure 4. The rail station model shown above is composed of 'Inline' nodes, described subsequently, as well.



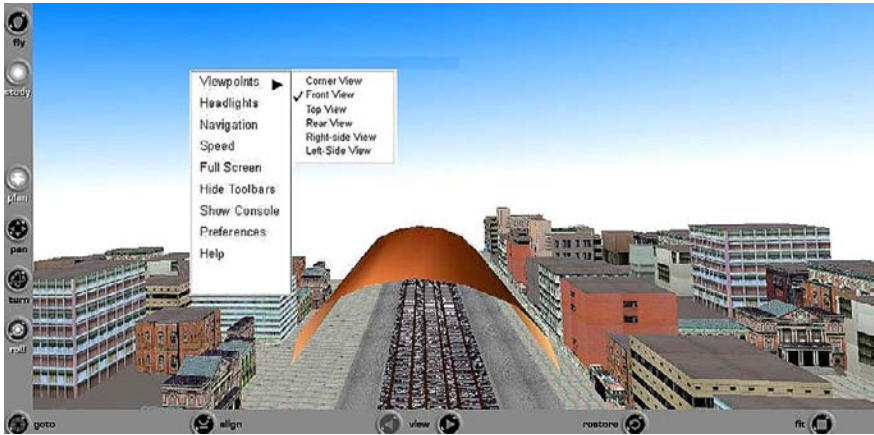


Figure 4: VRML world with multiple viewports.

<pre> Anchor [# LINKING to VB form children [DEF Platform1 Shape { ... }] url " C:\RailStation-3D\Platform1.frm" description " Database-Platform1" parameter ["target= frame_Platform1"] </pre>	<pre> Inline { url [C:\compartment1.wrl] bboxCenter 0 0 0 bboxSize -1 -1 -1 } </pre>

Figure 5: Anchor node and Inline node functionalities.

Once the 3D component of the model generation is done, any descriptive information or annotation can be associated with the corresponding object. This may include any type of information as the application demands. For instance, an application might demand the linking of GPS positional information and another application might require facilitating user-input by means of visual basic forms. This is done by means of the 'Anchor Node' construct shown above. Similar to the Anchor node functionality, another node that is used in the construction of databases or 3D worlds is an Inline node illustrated above. A VRML scene can be divided into a set of files. This not only simplifies the design of the VR world, but also facilitates reusing parts of the world already built in one scene in many other worlds. In the case of the railway station, a set of shapes that draw a train or compartment are grouped together and stored under a single node. This group node can be given a specific name, say 'compartment1', and reused later in another world. The Inline node permits the specification of a URL from where a specific file or data can be retrieved and reused in another file [7].

4 Dynamic models: simulation systems for training

This section focuses on the development of dynamic simulation models that can be used in the process of driver-training. Simulation is also a very powerful tool for collaboration, particularly with the practically unlimited connectivity and access provided by WWW. Manufacturing, assembling, and testing a prototype locomotive engine or rail car is a laborious and time-consuming task indeed. It is also an expensive process. We see that there are several aspects to be considered



<pre>TimeSensor { enabled TRUE startTime 0.0 stopTime 0.0 loop FALSE cycleInterval 1 }</pre>	<pre>DEF ClockTimer TimeSensor {...} DEF RailPathPositionInterpolator {...} ROUTE Timer.fraction_changed TO RailPath.set_fraction ROUTE RailPath.value_changed TO Train.set_translation</pre>

Figure 7: TimeSensor node and the ROUTE function for train movement.

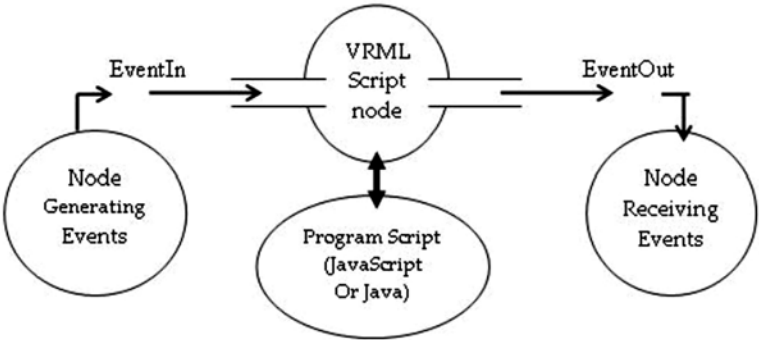


Figure 8: Flow of events among the nodes and the SCRIPT program.

4.1 Designing dynamic models using event-driven approach

Having seen the advantages of using 3D simulations, the discussion now focuses on the actual animation framework within virtual worlds. As mentioned in the previous section, using JavaScript or Java in conjunction with VRML 2.0 provides animation and interaction capabilities to the VR world.

The Script node in VRML calls upon a JavaScript linked with the node to function within the browser. While this refers to the internal control of the scene from within a node in the VR code, the EAI (External Authoring Interface) allows manipulating the scene externally. Movements within a VR world are brought about by animating position, orientation, or scale of the coordinate systems in which the train objects exist. A clock is created by means of the TimeSensor node and the changes can be described using the PositionInterpolator and OrientationInterpolator nodes. In this case, the simulation depicts the movement of a train along rail paths between stations or along test routes. The path is a linear or a curvilinear path. The position along this linear or curvilinear path is specified using PositionInterpolator and OrientationInterpolator nodes.

The events must be transmitted from the TimeSensor to the PositionInterpolator and from the PositionInterpolator to the Transform node. The ROUTE statement is used to link event from one node to another node. The output event from PositionInterpolator, value_changed is routed to the translation field of the Transform node containing the 'Train' object. As

animation progresses, the clock continues to send changed time values (SFTIME) and correspondingly new positions are chosen and output. This serves as the input for the 'Train' object, thus causing the node's coordinate system to undergo corresponding transformation (translation, scaling, and/or rotation). As a result, the 'Train' movement or simulation is achieved (Figure 9).



Figure 9: Simulation training for drivers using virtual worlds.

5 Discussion and conclusion

The primary aim of this study was to explore the use of virtual reality tools and techniques in building sustainable railway transit systems. This study discussed the process of a prototype model generation using VRML and animating the scene using a SCRIPT node and a scripting language, JavaScript. Even though the conventional drawings and other 2D representations may provide an idea of the place being designed, these cannot substitute a 3D view. Present rail station design and locomotive modeling efforts are far more complex when compared to those a couple of decades ago. Hence, it is imminent 3D data be viewed in their 3D forms in order to gain a better insight. Also, during the design and planning process, it is common to come across a situation wherein there is a need to visualize a future scenario.

This study explained with examples and prototype VR models the design and construction of railway transit systems and also the use of such systems to provide training to drivers. This study is just a beginning in the efforts to utilize the immense potential of virtual worlds in railway transit designing and planning. Future research may focus on the design of innovative locomotive and rail car designs and the creation of advanced simulations that impart state-of-the-art training at extremely low cost without wasting any resources.



References

- [1] Doyle, S., Dodge, M., & Smith, A., (1998): The potential of web-based mapping and virtual reality technologies for modeling urban environments. *Computers, Environment and Urban Systems*, 22: 137 – 155.
- [2] Worboys, M. F. (1994): Object-oriented approaches to geo-referenced information, *International Journal of Geographical Information Systems*, 8(4), pp. 385-399.
- [3] Section 4 Design guidelines Railway Design guidelines and criteria November 2005.
- [4] Michael Meillard, 1995. Reinventing the railway station, *Japan Railway and Transport Review* 1995.
- [5] Jepson W, Ligget R And Friedman S (1995), An Environment for Real time urban simulation, I 1995 Symposium On Interactive 3d Graphics, ACM.
- [6] Sheppard, S.R.J. 1989. *Visual Simulation: A User's Guide for Architects, Engineers, and Planners*. Van Nostrand Reinhold, New York.
- [7] Ames, L.A, Nadeau, D.R., Moreland, J.L., 1996, *VRML 2.0 Sourcebook*, John Wiley & Sons.



This page intentionally left blank

The network effects of railway investments

S. Hansen¹, A. Landex² & A. H. Kaas¹

¹*Atkins Danmark A/S, Denmark*

²*Centre for Traffic and Transport, Technical University of Denmark, Denmark*

Abstract

Network effects are when a change at one place in the railway network results in changes elsewhere in the network – maybe even far away from the original change. Railway investments have network effects, and therefore, this paper describes the network effects and how these network effects can be examined by queuing time. This paper gives examples of network effects and describes the importance of the size of the analysis area and the connections between trains in the railway network.

Keywords: network effects, queuing time, timetabling.

1 Introduction

The aim of many railway investment projects in Europe is to remove bottlenecks in the infrastructure, thereby making it possible to reduce travel times and increase the number of trains. The size of the benefits depends of the actual project involved. In any case, it is an important planning task to determine the future timetable. This makes it possible to determine the benefits for travellers in a standard cost-benefit analysis (CBA). The scope of this paper is to discuss the influence of the size of the analysis area on the calculated travel times.

Now, once a project is completed it is fairly straightforward to recognize that the project will also influence the future timetable outside the project area, thereby creating benefits outside the local area of the project. This is denoted “Network effects” in this paper. It is not possible to claim this as a general effect.

On a rural branch-line, where local trains connect to the national InterCity system, it is unlikely that any project will influence the national timetable. The benefits on the rural branch-line will be purely local, while removing a major bottleneck on the main-line network is likely to create changes on many



connecting lines, and maybe improve the timetable across the national network. In this case, a local analysis is not sufficient to capture all benefits.

1.1 Network effects in the Danish context

An example to illustrate the network effects is the Danish railway line between Aalborg and Frederikshavn, cf. figure 1. It is a single track line with a one-hour service. The travel time in one direction is 63 minutes and 66 minutes in the other direction. The speed on the line is increased from 120 km/h to 180 km/h.

This project can be evaluated locally. However, the traffic in the northern part of Denmark is not timetabled independently of the remaining network. The trains are part of the nationwide IC system and are therefore adapted to the arrival and departure times of the IC-trains at Aalborg (as well as the crossing possibilities in the northern part of Denmark).

If the crossing in the candidate timetable for the upgrading project is moved to obtain benefits locally, e.g. 10 minutes for one of the directions, it would result in nationwide changes. It is due to the fact that most regional trains have connection to and from IC-trains. A change in the northern part of Denmark will therefore influence the regional trains Copenhagen-Nykøbing F (in the southern part of Denmark) because of the connection at Ringsted cf. figure 1. This change may very well result in time benefits (or losses) at other lines of the network.



Figure 1: The Danish railway infrastructure.



Since network effects can influence the entire railway network, the extent of which are impossible to examine, it is necessary to delimit an area of analysis. Therefore, it is interesting to examine the importance of these network effects.

2 Methodical overview

To identify the network effects of the above mentioned increased speed project, a nationwide candidate timetable for one standard hour must be worked out. However, the nationwide timetable in Denmark depends on the train services to/from Germany and Sweden. To evaluate all the network effects it is therefore not enough to create a nationwide candidate timetable. It is necessary to include the trains to/from Germany and Sweden and thereby also the nationwide timetables of Germany and Sweden and so forth.

Network effects can be illustrated by queuing time. Queuing time is the difference in running time when comparing a single train on a line with a situation with many trains on the line. Queuing time on railway lines occurs when the traffic intensity is close to the capacity level due to e.g. mixed operation (slow and fast trains). When close to the capacity level, the operation speeds of fast trains must/will adapt to the slower trains cf. figure 2. This will increase the travel time for the trains that under free conditions could run at higher speeds [6].

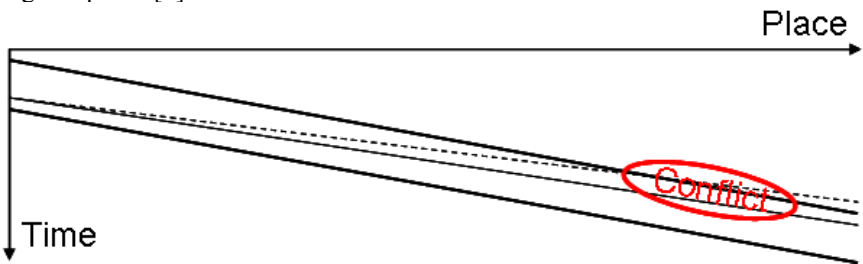


Figure 2: Extended running time (queuing time) due to other trains on the line [6].

To calculate the queuing time the Danish developed SCAN model (Strategic Capacity Analysis of Network) can be used (a similar function is found in the German tool UX-SIMU). SCAN is a computer tool for calculation of capacity in a railway network. In SCAN capacity is measured as average queuing time in a sample of candidate timetables for a given infrastructure alternative. The tool can be used in the strategic planning process where the exact infrastructure and timetable are not determined. Therefore the system is based on a structure where it is only necessary to know the plan of operation (i.e. the number of trains within each category), the infrastructure in a simple way and the main dynamics of the rolling stock [3].

2.1 The general problem of network effects

One of the problems you have to handle when calculating the consequences of a project is the size of the analysis area. When the analysis area is large, the risk of network effects is high. This is due to the fact that when a large analysis area is examined it is because of bigger changes in the infrastructure and/or timetables. Major changes in the infrastructure and/or timetables may influence many trains in the analysis area, and these trains may influence other trains outside the analysis area.

However, even smaller analysis areas may generate network effects. This is due to the way of planning the timetable in Denmark and many other countries. All train services can be placed in a hierarchy, cf. figure 3, where the train services placed in the top of the hierarchy is planned and timetabled before trains further down in the hierarchy.

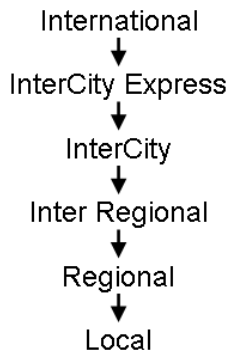


Figure 3: The hierarchy of the train service.

Even small changes in the timetable of a train in the upper level of the hierarchy may influence other trains further down in the hierarchy, because these trains are planned according to the train high up in the hierarchy. Since trains high up in the hierarchy often travel long distances, the changes for other train services can occur far away from the analysis area.

These changes in the railway network can occur far away from the analysis area is due to the network effects. Therefore, it is important to examine whether there are network effects. Nonetheless, the effects are normally only studied locally. It can be due to lack of resources, or because the network effects are uncertain (or insignificant), or because you only wish to evaluate the project locally, isolated from the remaining network.

2.2 Network effects on the railway line between Copenhagen and Ringsted

An example of a local study is found in the Copenhagen-Ringsted study where it was examined whether to extend the existing railway line or build an extra railway line in another layout. The capacity analysis for each of the alternatives only covers the actual line Copenhagen-Ringsted. For instance, the capacity



conditions at Copenhagen Central Station have not been included. The analysis itself consists of calculated queuing time for basis and 3 main alternatives as well as a variant of the new construction solution see figure 4. The analysis can be used to rank the alternatives examined, but it does not provide information about the conditions of the project alternative, when the capacity conditions on the adjacent lines are included.

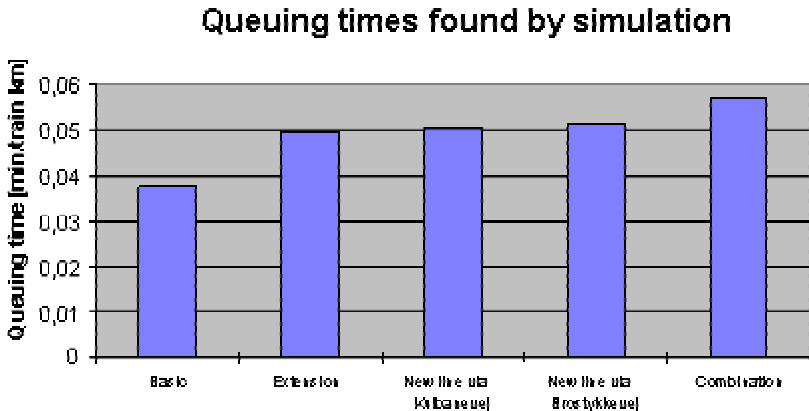


Figure 4: Queuing time analysis of the Copenhagen-Ringsted alternatives

However, when an infrastructure project is carried out, the traffic on the line or station in question is of course not planned independently of the remaining network. As the yearly timetable is planned at the network level, the timetable for the project area will also be influenced by this. The extent of this influence depends entirely on local conditions. In case of major improvements of the infrastructure on the main lines, it influences the timetable for the entire country (due to connections between long distance trains and regional trains).

2.3 Methodical results

The influence is bilateral, partly because a candidate timetable covering the project area may have derived effects at the network level, but the opposite can also be the case. If a candidate timetable is only worked out locally, it is not examined whether this timetable has conflicts outside the area. At worst, the candidate timetable in question gives rise to conflicts on the adjacent lines in the immediate area. Therefore, it can not be implemented in real life because of capacity conditions on the remaining network and it thus overestimates the benefits of the project.

The problem will be most significant, if the project is adjacent to lines with dense traffic/high capacity utilization. When preparing the timetables you normally start with the lines forming bottlenecks within the network, and if the



project is adjacent to such a bottleneck, the timetable for this line will determine the timetable for the project area.

A way to recognize the network effects is to work out a candidate timetable for a bigger part of the network than the project area. The size of the network to be included will depend on the project (and the analysis level). This method is used in the Netherlands where the network is more complex than in Denmark. Here a candidate timetable for one standard hour is worked out for the entire network. This is done by the Dutch tool DONS [5]. The timetable is then used to evaluate a project, so that it reflects both local effects and network effects.

It is a methodical problem that network effects cannot be generalized. Analyses of queuing time can be used for an initial evaluation of the network effects. By first carrying out an analysis of the queuing time locally and then an analysis including the adjacent lines the difference can indicate the problem. If the level of queuing time increases (queuing time per train-km) it is necessary to look at the network effects, but if the level decreases they can (probably) be ignored. This method is tested in the following section.

3 Network effects illustrated by queuing time

In this section the size of the network effects is examined for the Copenhagen-Ringsted example by means of the queuing time method. After the public hearing in 1998, a new examination of the project was initiated in 1999, including the development of a new traffic model. For the alternatives to be examined, a set of candidate timetables were prepared.

Three of these alternatives have been selected and used in this section. With regard to the infrastructure, the alternatives are as follows:

Table 1: Scenarios examined.

Scenario	Description of scenario
Basis	Infrastructure in Funen and Zealand as in 1999
Extending line	Two extra tracks between Copenhagen-Høje Taastrup and Roskilde-Ringsted and higher speed on the line
New Line	Two extra tracks between Copenhagen and Ringsted and higher speed on the line

The candidate timetables comprise the railway network in Zealand (except from the railway line between Elsinore and Copenhagen Airport) and Funen and include the line Copenhagen Central Station-Fredericia (just west of Funen), cf. figure 1 and figure 5.

With respect to this paper these timetables have been reduced to plans of operations by eliminating all arrival and departure times, and the queuing time is then simulated in the tool UX-SIMU. In this model, queuing time is calculated as the average of all "timetables". The result is minutes of queuing time per train-km.



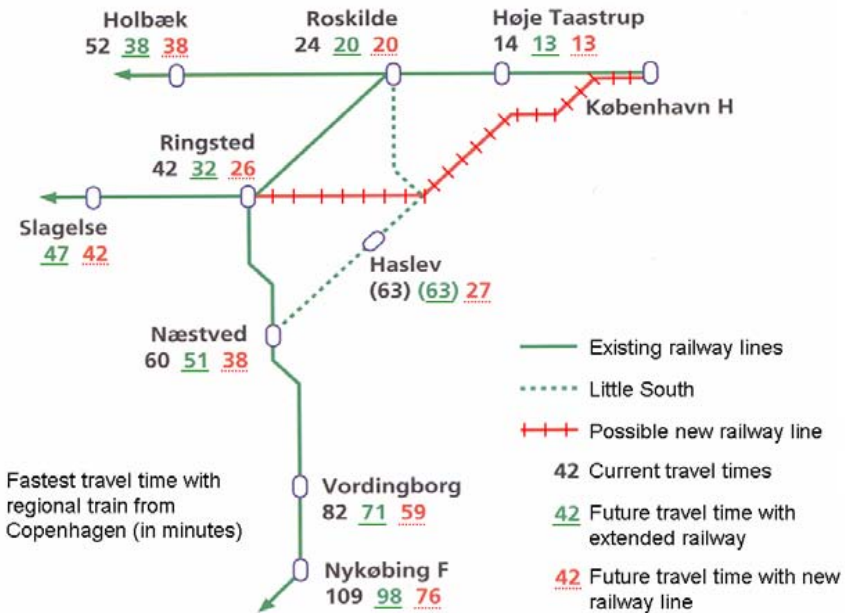


Figure 5: Fastest travel time with regional train from Copenhagen [4].

Two studies of network effects are carried out:

1. The importance of the size of the analysis area
2. The importance of including connections

In the first calculation the analysis area is reduced to include only the project area locally. The queuing time is calculated for the same trains, but only on this line, so that it is only the capacity conditions on the approximately 60 km long line that influence the queuing time.

In the other study the original analysis area is maintained, but now 2 connections are included in the calculation of the queuing time. A demand for connection will reduce the number of possible candidate timetables and give a more realistic calculation. Traditional analyses of queuing time do not cover this, and it is an example of an obvious calculation simplification. It is not necessarily clear which transfers there will be between the trains in a future scenario, it is however taken for granted that there will be connections somewhere in the network. The need for resources is of course smaller if the connections are ignored, and for capacity analyses, where several projects have to be ranked, it is probably acceptable.

4 The importance of the size of the analysis area

The importance of the analysis area has been illustrated by calculating the queuing time for the whole of East Denmark as opposed to only calculating the



queuing time for the line that is extended, i.e. the Copenhagen-Ringsted line. In East Denmark the capacity conditions Copenhagen-Fredericia (approx. 200 km) are included as well as the adjacent lines in Zealand where also single track lines are included.

The result is made up as average queuing time per train-km, as the 3 main alternatives do not include the same number of train departures. The result appears from table 2. The average queuing time for a train that covers a distance of 100 km is for instance 4.2 minutes in the basis scenario.

Table 2: Queuing time in the Copenhagen-Ringsted project.

Area	Basis	Extending line	New line
	(Queuing time in minutes per train-km)		
East Denmark	0.042	0.048	0.057
Copenhagen-Ringsted	0.033	0.046	0.033
Ratio of analysis area	79%	96%	58%

In both Basis and Full extension (New line) scenarios it is seen that the queuing time drops considerably, if it is only calculated locally as opposed to a bigger part of the network. An isolated local examination will therefore underestimate the queuing time when the project is seen in connection with the remaining net. In New line, with 4 tracks to Ringsted, it reflects that the bottlenecks are now found on lines further away than Copenhagen-Ringsted. In the extending line scenario the amount of queuing time is more or less the same, so in this case there is a balance between traffic and infrastructure.

The figures of queuing time indicate that the capacity conditions are underestimated when you only look at the effects locally. It can also be interpreted in connection with the timetables and the travel times indicated in these. If the candidate timetable only comprises the line locally, then the possible travel times are overestimated. When the remaining part of the network is included, it results in an increase of the queuing time. In other words, travel times, as a consequence of the project, will be longer than shown by a local calculation of the travel times Copenhagen-Ringsted. Once completed, locally the timetable will depend on the remaining network.

Furthermore, when comparing the queuing time for the entire network in the three alternatives it is seen that both extension alternatives have relatively higher traffic intensity than basis. It can be interpreted in such a way that the extensions are trafficked too heavily as compared to basis; i.e. the regularity will be relatively worse than at basis (i.e. the socio-economic benefits of the projects are overestimated).

5 The importance of connections

In many (Danish) cases analyses of queuing time do not take transfers into consideration. To examine the importance of this simplification, it has been



examined how much the above calculated queuing time for the whole of East Denmark is influenced, if a number of connections are taken into consideration.

The calculation has been made with 2 commonly occurring connections: At Ringsted between an IC-train and a regional train Copenhagen-Nykøbing F and at Roskilde between an IC-train and a regional train Copenhagen-Holbæk (cf. figure 1). The calculation has only been made for the extension solution, the new line scenario.

Table 3: Queuing time Copenhagen-Ringsted with connections.

Area	Basis	Extending line	New line
	(Queuing time in minutes per train-km)		
East Denmark	0.042	0.048	0.057
Including connections at Roskilde and Ringsted	-	-	0.061

It appears from table 3 that the amount of queuing time increases a little in the New line scenario. However, as only very few of the trains examined are subject to connection demands, it requires further studies to make a general statement.

On the other hand it can be concluded that network effects can have a big influence on the locally obtained travel times, and thereby on the calculated socio-economic benefit. As a principal rule, candidate timetables are worked out for an area that is larger than the project itself to include the effect of the capacity conditions on the adjacent lines. As a supplement to this a sensitivity analysis can be carried out by means of a local timetable. In a number of cases the local timetable will give shorter travel times and a better socio-economic profit and illustrate the effect of bottlenecks in the adjacent network.

6 Conclusions

This paper has presented network effects and a method to evaluate the network effects by examining the queuing time in the timetable. It has furthermore been shown that network effects are likely to occur in a railway system.

Changes in the timetable at one place of the railway network can lead to changes in the timetable for train services and/or railway lines far away. These changes on other train services and/or railway lines far away are due to network effects. Network effects are most likely to occur if changes in e.g. the infrastructure in a large analysis area are examined or if a train high up in the hierarchy of train services is influenced by the analysis.

References

- [1] Hansen, S., Large Transport Infrastructure Investments and their Strategic Impacts with a Special Focus on Enterprises, *PhD. thesis at Centre for Traffic and Transport, Technical University of Denmark*, 2004 (in Danish).



- [2] Kaas, A. H., Methods to calculate capacity of railways, *PhD. thesis at Department of Planning, Technical University of Denmark*, 1998 (in Danish).
- [3] Kaas, A. H., Developing and practical use of a capacity model for railway networks, *Proc. of the 2nd International Conference on Computer Applications in Railway Systems*, 1998
- [4] The Danish Ministry of Transport, Upgrading the main railway lines *Baneplansudvalget*, April 1997 (in Danish).
- [5] Odijk, M. A., Railway Timetable Generation, *PhD. thesis at Delft University of Technology*, 1998
- [6] Salling, K. B. & Landex, A., Computer based ex-ante evaluation of the planned railway line between Copenhagen and Ringsted by use of a new Decision Support System named COSIMA-DSS, *Proc. of the 10th International conference on Computers in Railways*, eds. J. Allan, C.A. Brebbia, A.F. Rumsey, G. Sciutto, S. Sone & C.J. Goodman, 2006



Modelling and simulation of the traffic management in a migration phase: example of “Ligne 1” of the Parisian subway

M. Ghantous-Mouawad^{1,2}, W. Schön¹, J.-L. Boulanger¹
& G. Churchill²

¹*Laboratoire Heudiasyc, Université de Technologie de Compiègne, Centre de recherches de Royallieu, France*

²*Régie Autonome des Transports Parisiens, RATP, France*

Abstract

This paper gives a closer look to the subject of the conversion of a conventional metro line into a fully automated operation without interrupting or disturbing the service. In the introduction, we present the Line 1 Automation project as one of the principal aspects of Paris metro's network modernization plan announced by the Régie Autonome des Transports Parisiens (RATP). Next we give an overview of the present line 1 and illustrate the decision key elements for choosing it as a target of a fully automatic operation. The following section describes the development of the line 1 model and shows the results made so far. Finally, we expose the conclusions of the present work and the outlooks of the line 1 automation project.

Keywords: Paris subway network, fully automated operation, safety, simulation.

1 Introduction

Three phases have marked the RATP history. The period 1900-1935 was the birth of the Paris subway. The period 1955-1975 has known the first modernization wave with the technology breakthrough of the operational control centers (OCC), the automatic driving, the controlled manual train driving (CMC) and with the new generation of the rolling stock. The period 2005-2020 shall see the second modernization wave. A 50 years cycle gives rhythm to these evolutions. So the new phase engages the Paris subway for the next decades.

The functional and technical bases of the new modernization program have been defined by the RATP on April 26th, 2002. Being oriented towards the new



technologies, the equipments modularity and their interchangeability, they (it) have permit to define the program known as OURAGAN (Offre Urbaine Renouvelée et Améliorée Gérée par un Automatisme Nouveau). The first implementation of OURAGAN shall take place on line 13 in 2007.

The modernization takes into account the accident report of Notre Dame de Lorette. On August 30th, 2002 at 13h21, the leading car of a train of the line 12 lies down on the track in the distance between the stations Saint Georges – Notre Dame de Lorette, slips of a 50 meters distance and ends up by colliding with the platform edge coping, face to a train stopped in the opposite track. This accident, whose origin was due to a significant excess of speed, has led the RATP to program the network equipment of a continuous speed control in less than 15 years instead of 30 initially foreseen.

The report of the functioning of line 14, a fully automatic line (Line 14 put in service in October 1998), whose quality of service is approved by the majority of the French citizens and the operators around the world has been also integrated and the eventuality of automating an existing line is stated.

It is within the framework of this policy that the RATP has announced in January 2003 the launch of a feasibility study of line 1 automation Churchill [1].

The main challenge of that project is not automation itself, whose feasibility studies have been carried out and demonstrated by different well known projects made all over the world, including RATP previous experience with SAET-Meteor, which is recognized as a technical and an economical success. Moreover, operation interruption of one of the most loaded Paris Metro lines being impossible, the main goal (which is a world's first) is to manage the migration phase, new automation equipments being installed during nights with the strong constraint to be able to resume normal operation a few hours later at the opening of the line. This task must be achieved without significant decrease of the quality of the service offered to the passengers and without any degradation of people safety, including operation staff. This latter concern is the cause of very specific problems during the transitory phase where cohabitation between manual and automated trains will be unavoidable.

In this context, RATP has developed a modeling and simulation software tool in order to be able to carefully analyze the different possible operation scenarios in the migration phase and therefore anticipating and assessing the subsequent risks. The aim of this work is to offer to decision makers a complete model of the line that allow to test different operation strategies during the phase where fully automatic trains and conventional manually driven trains will run together on the line, their performances and their capacity to adapt to technical, organizational and environmental disruptions. This model takes into account all the technical elements that are part of the control chain of trains movement's command, such as infrastructure, traction power supply, signaling, rolling stock, all in nominal and in all identified possible failure modes. The simulation is focused on the optimization of the operation, on the risk tolerance in cases of human or technical failure. Graphical user interfaces are developed for any potential user (engineers, trainers, and operators), the model being used as a validation but also as a training tool.



2 Line 1 overview

Line 1 presents the following characteristics:

- Length : 16.6 km
- Number of stations : 25 (23 underground stations)
- Number of trains : 52 (6 cars – 90 m, rubber tyres rolling stock/ third rail power supply)
- Headway : 105 sec
- Trip time : 76 min
- Capacity : 23.500 pas./hr/direction
- Commercial speed : 27.4 km/hr

One of the decision key elements for choosing line one to be equipped for fully automatic operation is the high request for adaptability and responsiveness of the transport offer.

An essential central axis of the common transport in the Ile de France region, Line 1 is extremely sensible to the demand variations because of its many connections and the areas that it crosses. The analysis of its traffic shows that in spite of a satisfactory average offer, line 1 does not respond to the travelers' requests and does not offer the expected service. We observe notably a particular overload during the rush hour, off-peaks hours, weekends and holidays

Its traffic is the most important in the subway network with an average of 725.000 trips per day and 207 million trips per year.

The screen doors, a complementary element of the automation avoid all the severe and serious travelers' accidents: falls on the tracks, suicides..., also frequent on line 1 and enhance the line regularity. Figure 1 represents the distribution of delays due to passengers based on 2003 data.

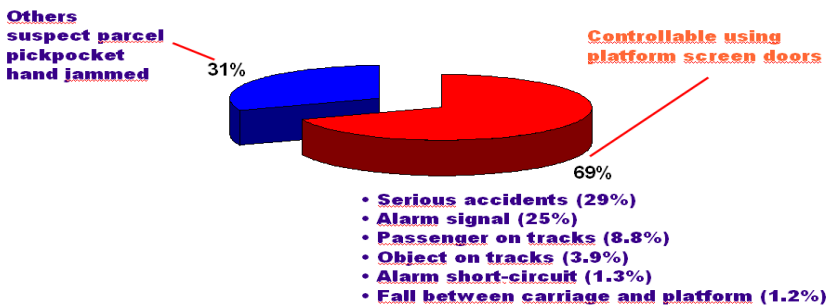


Figure 1: Distribution of delays due to passengers (last updated in 2003).

3 Overview of the present safety management of the traffic

The safety process is already identified for a conventional line such as Line 1 and for a driverless line such as Line 14. For each case of irregularity, accident, incident... specific regulations are to be followed and safety measured are to be



made according to the regulations in order to minimize and if possible to avoid the negative consequences. Each time an agent notice an incident or an irregularity on a train, on a fix installation or in a station that disrupts or risks to disrupt the operating, he must inform and warn immediately the operator in charge at the OCC with all the useful details.

The operator in charge at the OCC, informed of an incident or an irregularity, take the necessary measures. In case of a driverless train, he is also kept informed by the specific alarms of incidents, irregularities occurred in the fields managed and supervised by the SAET.

The agents as soon as they are informed of the incident announce their presence to the chief operating manager who coordinates their intervention. He gives all the useful indications to the drivers and the station agents in order to inform the travelers.

3.1 Definitions

3.1.1 Carrousel

The carrousel manages the movement of the trains in the parking zone. It is defined by an entry position, an exit one and by the intermediate places between the two extreme positions. When a train penetrates the carrousel from the entry position, it is moved automatically from a parking place to the next one in the carrousel until it reaches the first unoccupied place. This recent technique, already applied on line 14, and compared to the traditional train movement management in the garage zone, has a benefit in saving time on operational basis inside a terminal.

3.1.2 Priority table

The priority table is a table that indicates on daily basis of service the different positions of the garage existing in the terminal. It gives for each position a certain priority used to determine the sequence of the parking mission. The priority table allows defining the type of the trains. In this table and for each simulation, we define the initial occupation state of the places in the garage, the number of equipped trains and non equipped trains present in the terminal and the rest of line 1 and their directions. A train is assigned a unique identity number. This number is determined in the priority table.

3.1.3 SAET architecture

The SAET architecture represented in figure 2 is a recapture of the OURAGAN architecture. The SAET is composed of the following technical modules:

- Modules of sector treatment which make it possible to safely manage the trains running. They are composed of sets of two calculators DIGISAFE in hot redundancy.
- Onboard treatment modules which make it possible to command the rolling stock. They are also composed of sets of two calculators DIGISAFE in hot redundancy. (Digisafe is a safety architecture for critical numeric applications, based on the coded monoprocessor principle. Following that principle a single processor manipulates coded



variables and is therefore able to detect, on line and in real time, its possible errors (consecutive for example to hardware failures or perturbations from the environment) and in that case to act conveniently (going to a safe state by cutting off the power supply of the safe outputs, leading in general to stop the trains in the concerned zone.)

- Track negative detectors that detect a trespassing of a shunting signal in the stop position
- Input/Output modules allowing the SAET to interface with external systems.
- Data transmission system as a transmission vector between the different equipments of the SAET. It is composed of a double ring computer network and a radio transmission allowing the communications onboard/ground based on a product line used within the framework of the projects of New York, Barcelona and OURAGAN line 13
- The server of the maintenance help system
- The traction logic made out of relays of NS1 safety level (NS1 (Signalling Relay) is a design technique for safety relays, making it impossible to remain (consecutively to a failure) in the excited state (for example the return toward the desexcited state is ensured by gravity instead of a spring). The desexcited state is therefore designed to be the safe state (leading in general to stop the trains.)
- The OCC is composed of an operation room, a technical room and a training room that includes all the necessary hardware needed to simulate the traffic of the line and to replay real recorded situations.

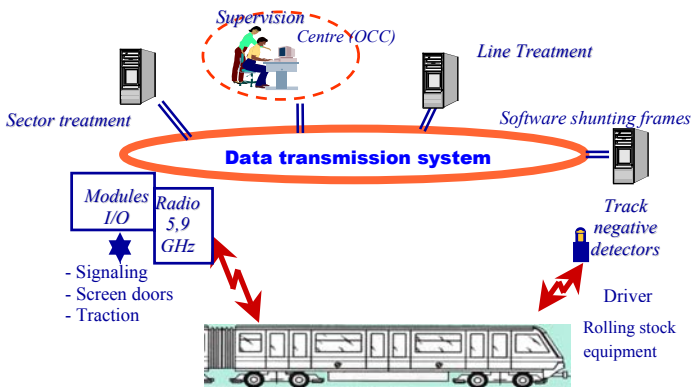


Figure 2: SAET architecture.

3.1.4 Space-time principle

A strong constraint of the traffic management is taken into account in the present Line 1 model: an equipped train is not allowed to run in a zone where agents are present on the tracks. Conversely, the access on secondary tracks is denied to drivers when an equipped train is detected in the related zone. In other words, an equipped train and a non equipped train cannot be present in the same place at

the same time. On the other hand, they can occupy a same place but at different moments. This principle is known as space-time principle.

3.1.5 Timetable

The timetable is the daily plan of missions by which the train service is required to operate which can be selected prior to the start the operating day.

3.2 Overview of the trains running management in case of a conventional line

The line operating is normally checked and supervised by operators known as “chief operating managers” at the operational control centre (OCC).

The monitoring and the responsibility of the trains running as well as the definition and the implementation of the measures to make in case of incidents rest on:

- the chief operating manager and if necessary on the manager(s) of the terminals in case of a line managed by a modernized OCC
- in case of the other lines: on the chief operating manager for the line and on the chief departure manager for the terminals.

The coordination is managed by the chief operating manager.

In case of an incident and in the case of a line managed by a modernized OCC, the chief operating manager, if necessary the terminal manager(s) coordinates the terminals agents’ tasks and decides of the measures to take in order to ensure an optimal service and to absorb the delays comparing to the normal basic schedule.

In the other cases, the chief departure managers coordinate the terminals agents’ tasks. The chief departure manager, in constant communication with his homologue in the opposite terminal, decides of the measures to take in order to ensure an optimal service and to absorb the delays with regard to the original schedule.

In case of an incident, the chief operating or departure manager can, if necessary cancel one or many trains or modify a mission.

3.3 Overview of the trains running management in case of a driverless line

The line operating is normally controlled and checked by the SAET equipments and supervised by operators known as “OCC chief supervisors of operation” at the operational control centre (OCC).

The monitoring and the responsibility of the line operating, the trains running, the supervision of SAET equipment and the information to the travelers as well as the definition and the implementation of the measures to make in case of incidents rest on the OCC chief supervisors of operation.

The coordination is managed by the OCC chief supervisor of operation that fulfils the functions of the incident chief.

He decides also of the measures to take in order to ensure a booked stop and an optimal service and to absorb the delays comparing to the normal basic schedule.



In case of incident (obstacle detection ...) interrupting the line examination, the OCC chief supervisor of operation asks for a driver intervention on the train board.

The OCC chief supervisor of operation, according to the information given by the driver:

- Either authorizes the automatic train operation for the rest of the trip while the driver stays onboard till the end of the trip.
- Or orders the manual train driving for the rest of the trip.

When the deterioration is perceptible, the OCC chief supervisor of operation asks the intervention of a qualified agent from the concerned technical service to check the concerned area.

The line examination being done and the potential incidents being solved, the line operating is then authorized.

The regular trains are put in service, automatically by the SAET, in accordance with the daily operating program.

A train needs a mission in order to move. The mission defines the trip and the tasks to be done by a train.

The management of these missions is ensured by the SAET from the OCC. These missions are normally chained.

4 Results

The specific safety regulations are not yet defined for the transitory phase. Therefore RATP is developing a software tool that allows the simulation of the traffic of line 1 in the present, the transitory and the target phase, in order to determine the safety measures for this phase without disturbing the service.

The results made so far are the development of the train based on the specifications of [4] and of the traffic of the two terminals of Line 1: Grande Arche de la Défense and Château de Vincennes based on the specifications given by [2, 3, 5].

4.1 Train development

4.1.1 Brief description

The main functions of a train are the traction, the braking and the service to passengers. It is composed of six cars, four of them are central driving cars (N) and the other two are trailers (S). In case of an ET, a trailer has a control panel. In the other case, it has a control cab. The train model developed is composed of the dynamic model of the MP89 coupled with the feedback control loop. This loop helps the train to calculate his position, velocity and acceleration while taking into account the values of these outputs at the previous moment. We calculate the train position by simple integration of the velocity while taking into consideration the traffic direction. The dynamic model of the MP89 is based on the resolution of the differential equation of the train movement for given characteristics of the line and the rolling stock, by analyzing the forces that appear at the rail-wheel contact. Consequently, the speed and the acceleration of



the train are known. By simple integration the speed and by taking into account the traffic direction, we determine the train position. The train feedback performances are identical for the future driverless train of line 1 (MP05) and the present conventional train (MP89), except the driver reaction times in case of a MP89. It is taken into account when the driver changes his control cab in order to change the train direction (30 seconds approximately) and it is higher than the reaction time of a MP05 (2 to 3 seconds). An initialization phase is launched at the beginning of each simulation. It provides to the train the needed input parameters as well as the line 1 characteristics such as the declivity of the two main tracks and the speed limits implemented throughout the line in the two directions. These speed limits are not identical for the MP05 and the MP89. The train load (from 0 to 10 passengers/m²) is indicated also before each simulation. It characterizes the travellers flow in the terminal for a given schedule. The user will define the maximum capacity of the travellers in a menu. In the present train model, the defined capacity is 722 travellers.

4.2 Development of the traffic

Figure 3 gives an overall architecture of the terminal model and his environment. It is a structured architecture composed of six main blocks.

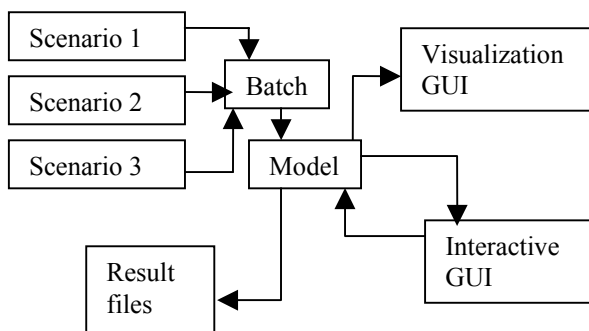


Figure 3: Terminal model structure.

Model is the central unit. Coded in Matlab/Simulink, it is responsible of all the basic operations. At the end of each simulation, two result files are generated. The first contains all the events already appeared in the GUI. The second gives more detailed information about the missions, the routes management and presents a summary of the delays of each train. The simulation reports in real time the position of the selected train, its type; its kilometer point, its velocity, its mission and the value of the delay. The visualization GUI displays also the significant events that happen during the simulation such as delays, train blocking situations, perturbations created, simulation hour and zone violation. This last point is based on the space-time principle defined above. The interactive graphical user interface offers the possibility to inject perturbations in



considering the rest of the line as a black box. Figure 4 is a visualization of Château de Vincennes animated by the trains running.

This model is currently being completed by inserting the rest of the stations. It will also take into consideration all the future improvements and modifications made on line 1 infrastructure and on the rolling stock.

6 Outlooks

In 2011, line 1 shall be entirely automated

- July 2007: start of the reinforcement work of the platform edge coping
- January 2008: start of the installation of the screen doors
- May 2008: delivery of the first MP05 to the centre of railway tests in Valenciennes
- July 2008: delivery of the first train for commissioning
- January 2009: Commissioning of the OCC version manual train driving
- July 2009: Commissioning of the OCC version train manually driven/driverless train
- August 2009: Commissioning of the first driverless train and start of the cohabitation period between driverless trains and trains manually driven running on the same line.

With a delivery rate around 2 to 3 trains per month, line 1 shall be entirely and completely an automated operation in February 2011. The RATP then shall be among the firsts in the world who has carried out and has accomplished the conversion of a conventional subway line into an automated operation without traffic interruption.

References

- [1] Churchill, "L'opération d'automatisation de la ligne 1 du métro de Paris", RATP, Paris, France, December 2005.
- [2] Département Métro, US FDC, Réglementation Ferroviaire et Mouvement des trains, CONSIGNE DE SECURITE FERROVIAIRE 01-001, "METRO LIGNE 1, GARE DE CHATEAU DE VINCENNES", RATP, Paris, France, March 1995.
- [3] Département MTS, US FDC, Réglementation, CONSIGNE DE SECURITE FERROVIAIRE 01-002, "MTS LIGNE 1, GARES DE PORTE MAILLOT, ESPLANADE DE LA DEFENSE, LA DEFENSE", RATP, Paris, France, January 2001.
- [4] ET 4-10460 G, "Spécification fonctionnelle MP89", RATP, Paris, France.
- [5] IEF-SR 04-5169, "Caractéristiques actuelles d'exploitation de la ligne 1", RATP, Paris, France, November 2004.



Computer based ex-ante evaluation of the planned railway line between Copenhagen and Ringsted by use of a Decision Support System named COSIMA-DSS

K. B. Salling & A. Landex

Centre for Traffic and Transport, Technical University of Denmark, Denmark

Abstract

This paper presents an appraisal study concerning the extension of the main railway line between Copenhagen and Ringsted by use of a software system named COSIMA-DSS. The modelling system is based upon a multi-methodological approach combining a cost-benefit analysis together with a multi-criteria analysis. One of the key impacts in cost-benefit analyses is the time benefit which depends on the timetabling. The paper describes some of the challenges related to the calculation of the time benefits and how the time benefits can be implemented in the decision support system. At the end of the paper, results from the calculations using COSIMA-DSS are presented and discussed on the basis of a newly developed Danish manual on transport appraisal.

Keywords: timetabling, queuing time, value of time, Decision Support System, cost-benefit analysis.

1 Introduction

Due to congestion it has been planned to increase the capacity of the main line between Copenhagen and Ringsted since the beginning of the 1990s. There are two different strategies to deal with the increase of traffic between Copenhagen and Ringsted, i.e. two more tracks along the existing railway line and a new railway line with a new layout cf. figure 1.



Due to increased traffic over the recent years, running times of the trains have been prolonged to obtain more capacity meaning longer travel times. The increase in time for the fast trains is significant, cf. figure i.e. on some travel relations the increase in time is up to 20% of the running time.

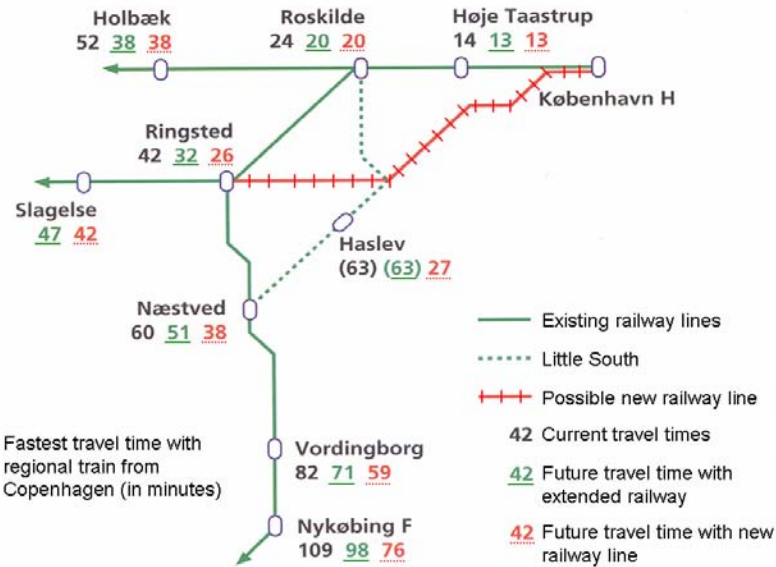


Figure 1: Fastest travel time with regional train from Copenhagen [11].

The decision whether to implement the new railway line or the extension is up to the decision-makers, in this case the Danish Government. However, by building a so-called decision support system (DSS), decision-makers can get “assistance” in making the best and most profitable choice from a societal point of view. It is necessary to stress that a DSS is not a correct and final answer to the problem it is merely assistance to the decision-makers.

2 Time benefits

In 2003 the Danish Ministry of Transport developed a set of guidelines on evaluating transport infrastructure projects [12]. When evaluating infrastructure projects using the Danish manual, the main impact group is the time benefits which in the evaluation of roads can make up a share in the range of 70-90% of the total benefits [7]. Therefore, it is important to make the correct assessment of the time benefits hence using the same guidelines in all transport projects for comparisons.

In schedule-based public transport systems, the time benefits can be divided into several elements – e.g. waiting time, transfer times, time spent in the vehicle and delayed time. Passengers consider each of these time elements differently



and several studies have therefore been carried out to evaluate these [1]. For instance it feels worse to be delayed than to spend time in the vehicle – therefore, delayed time has a higher time value than ordinary travel time in the cost-benefit analysis (CBA).

The present timetable for the railway line between Copenhagen and Ringsted has been developed, adjusted and improved over time. In this sense it is difficult and time consuming to plan an optimal timetable for a new situation with an alternative infrastructure. It is, however, important to plan the best possible timetable to achieve as much time benefit for the passengers and freight as possible. In addition to the previously mentioned time aspects, two other key elements should be taken into account, namely the queuing time between trains and the timetabling. These impacts are highly relevant when evaluating public transport systems; however, they are not applied in the manual.

2.1 Queuing time

The queuing time is a time element in the appraisal study worth noticing. Queuing time on railway lines occurs when the traffic intensity is close to the maximum capacity level due to e.g. mixed operation (slow and fast trains). When close to the capacity level, the operation speeds of the fast trains must/will adapt to the slower trains cf. figure 2. This will increase the travel time for trains that under free conditions could run at higher speeds.

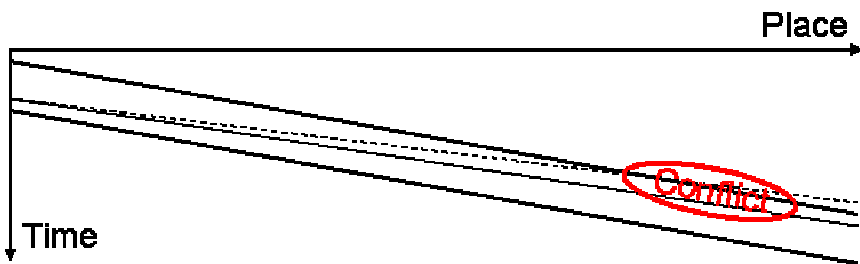


Figure 2: Extended running time (queuing time) due to other trains on the railway line.

The queuing time for the fast trains on the railway line between Copenhagen and Ringsted are significant, cf. figure 1. For some travel relations the queuing time is up to 20% of the running time.

Traditionally the queuing time is calculated as ordinary travel time for the passengers. However, it can be argued that the travel time should be divided into minimum travel time (including relevant time supplements) and queuing time. Furthermore, the queuing time should be assessed as a delay since the train and thereby the passengers are delayed due to other trains, even though it is a “scheduled delay”. Whether the queuing time is calculated as ordinary travel time or delayed time, it has a great impact on the result, since normal delayed time are weighted twice as much as the ordinary travel time (for commuters 59



DKK vs. 118 DKK [12]). The queuing time is an important factor particularly when making the timetables as illustrated in figure 2.

2.2 Timetabling

Calculating the time benefits only by use of the traditional CBA can result in the paradox that a well-planned timetable gives a worse societal impact than a sloppily planned timetable, when for example an extension of a railway line is proposed. This paradox is due to the lower socio-economic cost of travelling when the queuing time is considered as ordinary travel time (Scenario 1) instead of delayed time (Scenario 2) which will occur when the queuing time is not taken into account, cf. table 1. In the future scenario (Scenario 3) a time cost per passenger is evaluated to be 1,188 DKK. Comparing the future scenario with the two basis scenarios the sloppily planned timetable actually performs better than the well planned timetable from a societal point of view.

Table 1: Traditional calculation of socio-economic time costs.

	Scenario 1	Scenario 2	Scenario 3
Minimum running time (min)	32	32	32
Queuing time (min)	10	0	0
Travel time (min)	42	32	32
Delay (min)	0	10	0
Time costs per passenger (DKK)	2,478	3,068	1,888
Difference future vs. basis	590	1.180	-

Converting some of the mentioned aspects into larger scale is always a great challenge especially as the infrastructure or timetabling is changed and often not known. The next section introduces the time aspect as it is calculated in a large scale network.

2.3 Time benefits in large scale networks

The time benefits can relatively easily be calculated for the existing timetable on the existing infrastructure e.g. table 1. It is, however, much more complex to calculate the time benefits when the infrastructure and/or the timetable is changed. The increased complexity of calculating the time benefits is due to the uncertainty on how the infrastructure and/or the timetable will look in the future. Often more than one scenario is worked out and thereby also different candidate timetables. Since it takes a long time to create a timetable, the timetable is often only worked out for an analysis area. However, it is necessary to take the adjacent lines into account too because of the limitations of these line sections – the so-called network effects [3, 4, 6] If the network effects are not included, only the local time benefits are examined. When the network effects are examined benefits from the rest of the railway network are included in the analysis.



If a candidate timetable is only worked out for the analysis area it is not possible to calculate the time benefits precisely. One of the most difficult time effects to calculate is the queuing time. Previous studies on the queuing time [3, 4] have shown that the influence of the network effects can be significant for the queuing time.

The importance of the analysis area, and thereby the influence of network effects, can be illustrated by calculating the queuing time for both the whole of East Denmark and the analysis area only. The results are then made up as average queuing time per train km, as the 3 main alternatives do not include the same number of train departures. The results appear from table 2.

Table 2: Queuing time in the Copenhagen-Ringsted project [3, 4].

Area	Basis	Extended line	New line
	(Queuing time in minutes per train km)		
Whole East Denmark	0.042	0.048	0.057
Only the analysis area	0.033	0.046	0.033
Analysis area in proportion to East DK	79%	96%	58%

In both the basis and in the new line scenarios it is seen that the queuing time drops considerably, if it is only calculated locally as opposed to a bigger part of the network. An isolated local examination will therefore underestimate the queuing time when the project is seen in connection with the rest of the railway network.

In the extended line scenario with 4 tracks between Copenhagen and Ringsted the bottlenecks are now found on lines further away from the analysis area. This also reflects the amount of queuing time which is more or less the same – 96% of the queuing time per train km was found in the analysis area. In the extended line scenario there is therefore balance between the traffic and the infrastructure.

The queuing time analysis shown in table 2 indicates that the capacity conditions are underestimated when effects are only analysed locally. Furthermore, the results of the queuing time analysis shown in table 2 shows that it is important to include not only the analysis area but also the adjacent railway lines in the analysis of timetable scenarios since the queuing time otherwise might be underestimated.

Evaluating the Copenhagen-Ringsted project is especially difficult due to the previously mentioned time aspects. Further studies are necessary, especially when it comes to more widely applied effects e.g. network effects, queuing time and timetabling. These impacts are not yet applied in the Danish manual hence they are defined as strategic impacts. It is proposed to apply a multi-methodology approach [8] both combining the conventional cost-benefit analysis (CBA) approach with a more widely multi-criteria analysis (MCA) approach. The first step is to see the problem as a system and hereby model the various strategic elements within the system in a more qualified way – this leads to a systematic approach/assessment e.g. the multi-methodology approach.



The following section brings an overview of the applied multi-methodology approach together with the core impacts applied in this paper. Furthermore, the strategic impacts will be discussed as they are considered relevant for the Copenhagen-Ringsted railway line. Finally, all the effects are applied and evaluated in a socio-economic analysis.

3 The appraisal framework – COSIMA-DSS

The COSIMA-DSS (COMposite Model of Assessment-Decision Support System) model consists of two different modules brought together in the main module developed in Microsoft Excel [9]. The system shown in figure 3 gives a brief overview of the module structure of the model.

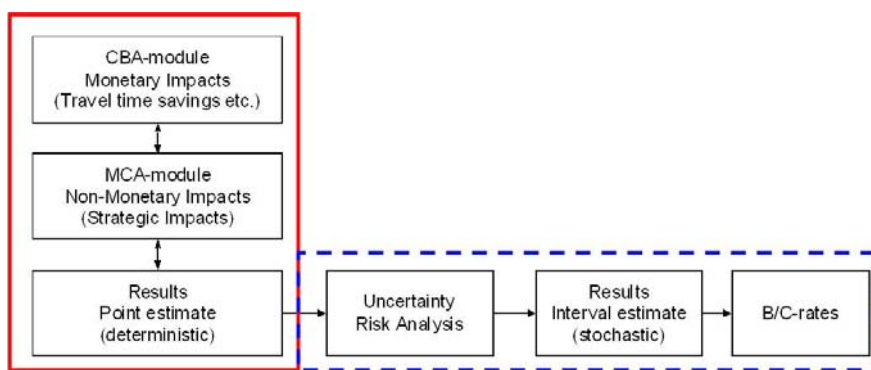


Figure 3: Main structure of the COSIMA-DSS model.

The left box consists of a so-called Value-Analysis module combining the traditional cost-benefit analysis (CBA) with the wider multi-criteria analysis (MCA). As shown in figure 3 the CBA module is of monetary character whilst the MCA module is of non-monetary character, in this case consisting of the previously mentioned queuing time and other strategic impacts as explained earlier. These different impacts (CBA and MCA) are combined in an Excel based model named COSIMA, determining a point estimate or rate of return, in this case by a benefit-cost ratio. The second box (on the right) is defined as the stochastic part that deals with the uncertainty present in all project appraisals. The uncertainty handling is performed by a Monte Carlo simulation which facilitates a complex analysis of the importance of uncertainty regarding some key input parameters. Furthermore, the project appraisal will have a more transparent perspective showing the degree of uncertainty for each element in the analysis ending up with an interval estimate as probability distributions [5]. In the following the deterministic box is applied to the Copenhagen-Ringsted framework. Recently, the stochastic modelling framework is not yet applied; however, in the future studies the stochastic calculations will be developed and implemented.

3.1 The CBA module

The basic principle in the COSIMA framework is the cost-benefit analysis (CBA) which is based on the assumption that society's welfare is measured by the preferences of each individual and their presumed value [1]. CBA is traditionally used in Danish appraisal studies when it comes to infrastructure investments, however, in the public transport sector appraisal methodology is not widely applied. This is due to the complexities involved in the determination of the different impact groups and "actors". Traditionally, the main input in a socio-economic analysis is the travel time savings, in evaluation schemes towards a railway line, this is only partly the case as several actors are involved, both operators and providers having to benefit from a new infrastructure investment. Furthermore, the strategic impacts such as the queuing time are important in order to make the overall performance of a railway network investment feasible as described in a previous section.

In the case of the Copenhagen-Ringsted railway line it is clear that by creating a new line or extending the existing line, the travel time will decrease meaning that the users will benefit from shorter travel times. Hopefully the operators will gain from an increase in travellers resulting in higher revenue and the providers will benefit from more travellers resulting in higher taxes etc. In figure 4 are the different impact groups together with their corresponding "actors" illustrated together with the benefit and cost groups.

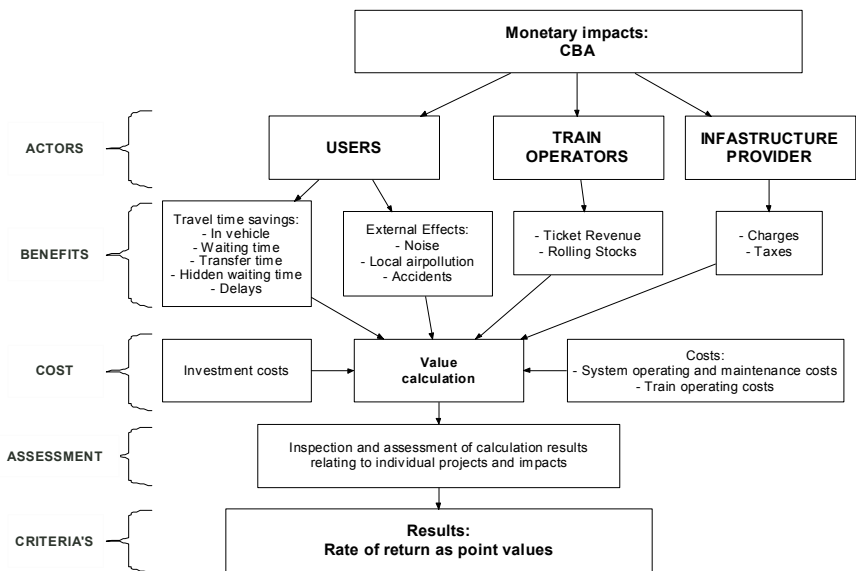


Figure 4: Flowchart on the different monetary impacts applied in the COSIMA-DSS model.

In the appraisal phase of the study, one general objective for the transport sector is to strengthen the competitiveness of the public transport as compared to



the road sector. This should be done via different secondary objectives such as high service frequencies, low travel times, high service reliability, high comfort and good transfer possibilities to other transport modes (intermodality). These different groups are in some sense incorporated within the travel time savings, however, the CBA requires that all relevant impacts of the project are assigned a monetary value. In the case where different time impacts need to be taken into account these should be considered as well in the decision process. Various impacts considered of a strategic nature – e.g. long term environmental impacts or economic impacts can be difficult to quantify and thereby apply a unit price. In this sense a multi-criteria analysis (MCA) can help the decision-makers to apply these more strategic impacts in the evaluation scheme. When combining the CBA and MCA, a multi methodology assessment of the railway line is possible, in this case resulting in a benefit-cost ratio.

3.2 The MCA module

The multi-criteria analysis makes use of the well-documented AHP (Analytical Hierarchy Process) approach to pair-wise compare the various impacts [2, 10]. Applying the pair-wise comparison makes it possible to assign a monetary unit to the MCA impacts even when quantitative ratings are unavailable. The different alternatives are assigned a score for each MCA impact and then the three MCA impact are compared in a software system named Criterium Decision Plus (CDP) by the AHP procedure [5], see figure 5.

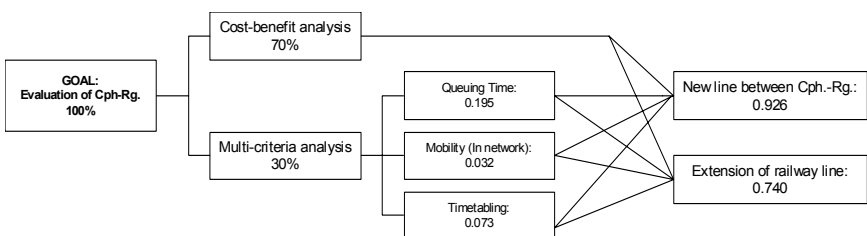


Figure 5: Model rates for the AHP procedure within CDP.

The output from the CDP is a set of normalized ratings of all the impacts which are transferred back to the CBA-module. The MCA impacts are then assigned a monetary value based on the comparison made in CDP. In this preliminary appraisal scheme it is assumed that the MCA counts for 30% and the CBA 70% of all evaluated impacts. The strength of the pair-wise comparison is that it allows or incorporates the decision-makers in the overall decision process. In discussion with the decision-makers it has been decided that three non-monetary effects should be taken into account, namely the queuing time, the timetabling and the mobility concerning the large-scale network cf. figure 5. The procedure is then to make the evaluation pair-wise considering all possible combinations between the three impacts. Finally, the two outer right boxes of the figure shows the preliminary ratings of the railway line – in this case it is the new line alternative which is merely better than the extension alternative.



3.3 Deterministic results

By use of the methodology developed by the Danish Ministry of Transport [12], the B/C-ratios for the conventional CBA have been calculated as the bottom column in figure 6. It is clear that the ratio for both alternatives is below 1.0 which means that none of the proposed alternatives are socio-economically feasible. However, as a result of the previous weighting schemes new calculations are performed to achieve total rate of returns (TRR) comprising both CBA and MCA.

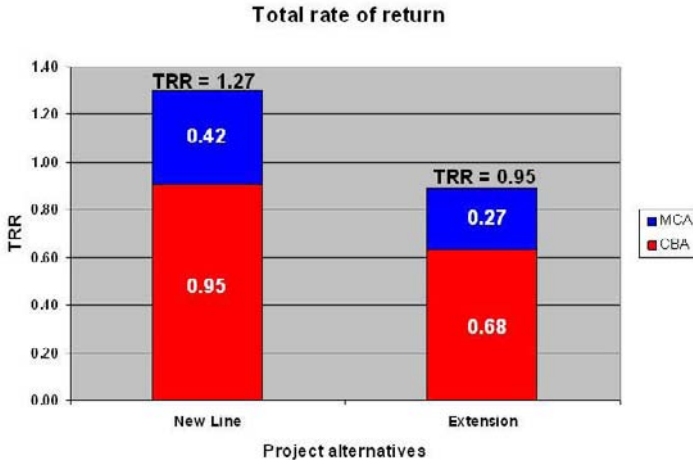


Figure 6: Graphic representation of a COSIMA-DSS calculation.

The COSIMA-DSS evaluation of the Copenhagen-Ringsted railway line shows that by assessing only two different alternatives the new line situation gives the best overall performance. Further development of the modelling scheme will be implemented as more alternatives are suggested.

Traditionally, evaluation methodology consists only of a cost-benefit analysis hence the bottom part of the columns shown in figure 5. However, none of the two ratios receives a positive rate of return which further concludes that the justification of implementing multi-criteria analysis is in order.

4 Conclusions

This paper has presented some of the challenges related to the calculation of time benefits on railway lines for socio economic analysis. It has been shown that the scheduled travel time for the passengers can be divided into minimum running time (including relevant supplements) and queuing time.

Furthermore, the paper has shown that the queuing time is difficult to calculate for future scenarios due to unknown timetables and network effects on large scale networks.

To deal with the difficulties of calculating time benefits the paper has presented a newly developed decision support system, COSIMA-DSS which



aims at assisting decision-makers in the appraisal of transport infrastructure project investments. The variety of different features embedded within a CBA and MCA approach makes it particularly useful for addressing complex transportation decision problems. COSIMA-DSS gives the decision-makers a set of tools relevant for planning and assessment of project proposals where a conventional CBA will be too narrow a methodological approach. Further model work and more comprehensive case studies will seek to demonstrate and validate the COSIMA-DSS approach.

References

- [1] Gissel, S., Decision aid methods in rail infrastructure planning, *PhD. thesis at Department of Planning, Technical University of Denmark*, 1999.
- [2] Goodwin, P. & Wright, G., Decision Analysis for Management Judgment. 2nd edition, pp. 387-399, *John Wiley & Sons Ltd., Chichester, England*, 1998.
- [3] Hansen, S., Landex, A. & A.H. Kaas, The Network effects of Railway Investments, *Proc. of the 10th International conference on Computers in Railways*, eds. J. Allan, C.A. Brebbia, A.F. Rumsey, G. Sciutto, S. Sone & C.J. Goodman, 2006.
- [4] Hansen, S., Large Transport Infrastructure Investments and their Strategic Impacts with a Special Focus on Enterprises, *PhD. thesis at Centre for Traffic and Transport, Technical University of Denmark*, 2004 (in Danish).
- [5] Infoharvest, Inc., Criterium Decision Plus – User's guide. Version 3, Infoharvest Inc., 1999, Seattle USA.
- [6] Landex, A., Kaas, A.H., Schittenhelm, B. & Schneider-Tilli, J., Practical use of the UIC 406 capacity leaflet by including timetable tools in the investigations, *Proc. of the 10th International conference on Computers in Railways*, eds. J. Allan, C.A. Brebbia, A.F. Rumsey, G. Sciutto, S. Sone & C.J. Goodman, 2006.
- [7] Leleur, S., Road Infrastructure Planning – A Decision-Oriented Approach, 2nd Edition, *Polyteknisk Forlag, Lyngby, Denmark*, 2000.
- [8] Leleur, S., Systemic Planning – Principles and Methodology for Planning in a Complex World, *Polyteknisk Forlag, Lyngby, Denmark*, 2004.
- [9] Leleur, S., Holvad, T., Salling, K. B. and Jensen, A. V., Development of the CLG-DSS Evaluation Model. Midterm Report No. 1 – April 2004, *Centre for Traffic and Transport – The Technical University of Denmark*, 2004.
- [10] Lootsma, F. A., Multi-criteria decision analysis via ratio and difference judgement, *Kluwer Academic Publishers*, 1999.
- [11] The Danish Ministry of Transport, Upgrading the main railway lines *Baneplansudvalget*, April 1997 (in Danish).
- [12] The Danish Ministry of Transport, Manual concerning socio-economic analysis – applied methodology in the transport sector, December 2004 (in Danish).



Assessing rail transport network performance and reliability

R. Raicu & M. A. P. Taylor

Transport Systems Centre, University of South Australia, Australia

Abstract

Network performance in general - and in the computing and telecommunications areas in particular - is a topic of major interest. Much of this research is not in the public domain, but this is beginning to change as international research teams become aware of common problems and issues. In any event, rail systems have certain special characteristics which limit the transferability of general network methods to rail problems or which require additional work to develop relevant interpretations and implementations for use in railway systems analysis. The paper presents results of a research project aimed at developing decision support tools for assessing rail network performance and reliability.

Keywords: rail transport network, network reliability and vulnerability, congestion, delay.

1 Introduction

When analysing network performance an integrated approach considering network infrastructure, operation and transport demand is essential.

The capacity of the transportation network can be evaluated by various measures such as the travel times and the extent of congestion. Network flows are influenced by abnormal events that affect network characteristics and capacity, like disasters, accidents, construction or repair. Ideally, networks should be designed so as to cope with normal fluctuations by offering alternative paths, but planning for abnormal events is much more difficult.

In systems engineering, reliability may be defined as the degree of stability of the quality of service which a system normally offers. In the face of increasing user demands for high levels of service, system reliability is becoming



increasingly important in the planning, construction and operation of transportation networks.

2 Model rail systems as formal networks

The rail transport network is a complex system made of numerous elements with functions determined in the transport process. The description of the rail transport network cannot be limited just to its elements; similar to other complex systems, its structure is very important. The concepts of graph theory are used to formalise the structure of the rail transport network.

The transport infrastructure network can be presented as a graph (K, D) , with K nodes and D arcs. The nodes of the network are just those points from which diverge or to which converge at least three arcs, and in which it is possible to transfer from one arc to any of the other two. Both arcs and nodes are characterised by certain attributes which can only be defined in correlation with the characteristics of the transport means they were designed for. Each arc (paths of the network) can be made of one or more homogenous sections characterised by length, and maximum, respectively minimum speed the transport means can operate.

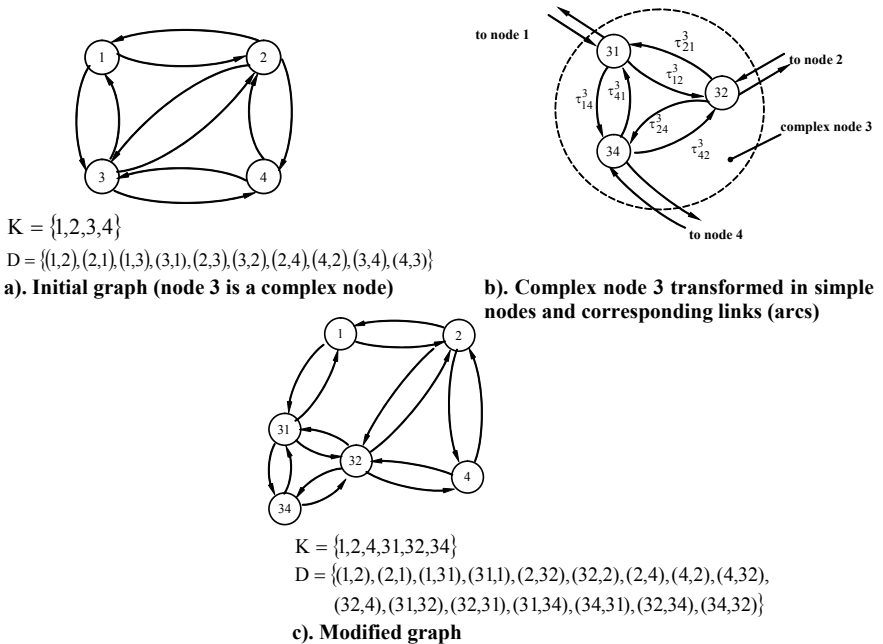


Figure 1: Obtaining the modified graph of the network.

The transit times cannot be neglected. These nodes, with their significant transit time are considered as being complex nodes. Figure 1 presents an example.



3 Particularities of rail transport networks

3.1 Basic principles

On the relatively simple graph of the network the set of paths (itineraries) which link any two nodes could be indicated. In describing the network the choice of nodes and arcs is essential. Different problems require different degrees of detail regarding the technical state of the network. Some examples of problems to address are: planning economic transport links, freight transport; planning links between urban concentrations, passengers transport; distribute traffic between transport nodes; determine traffic intensity on the sections of the network; planning investment works to increase transport capacity; evaluating resources use to accomplish the forecasted transport task; planning the technical and economical operating indicators of the transport mode and transport system as a whole.

3.2 Structure and contents of the information on the technical state of the rail network

The way the information is organised influences its efficient use for solving concrete problems. In selecting the organisation of information one has to bear in mind that certain parts of the information have to be easily accessible and partial modifications should not require the radical restructuring of the whole system. Thus a hierarchical structure of the information is preferred. For the rail network a three level hierarchical structure is recommended (see Figure 2).

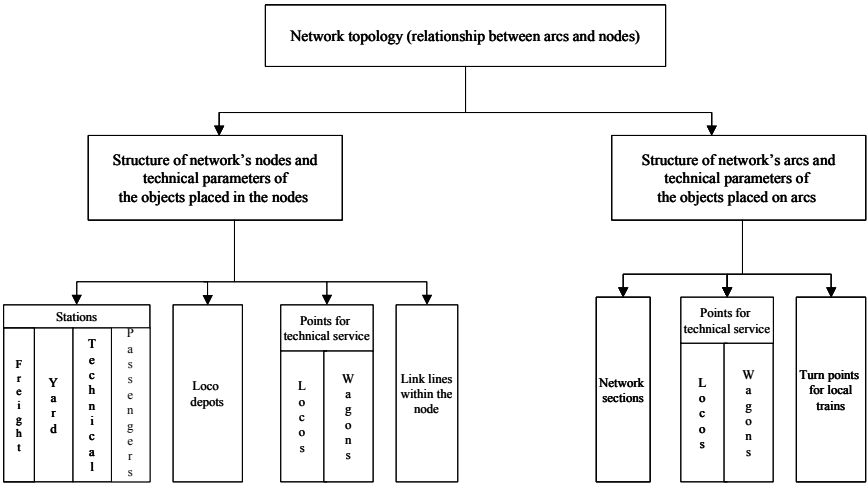


Figure 2: Structure of the information on rail network state.



The first level (superior) of information characterises the whole rail network and contains the list of arcs and nodes and the table of correspondence between them.

The second level of the hierarchic structure comprises information about the managed objects and the interactions between them. It refers to the objects in the nodes. Depending on the type of problem to solve different level of detail on the transportation process is necessary. The nodes of the network generally have a complex structure. One can identify in their structure two types of elements: objects and links between objects. In various transport models just a part of these objects is involved. For example, information about the location of the freight stations is necessary for the detailed evaluation of the material and financial expenses of the transportation operations. Technical stations and yards are included in the models of train formation on the network. The location of passenger stations and technical passenger stations is necessary to describe the passenger trains operation. Information on depots, locos and driver rosters are necessary to evaluate locos and drivers activity.

The third level of the hierarchic structure comprises information on the constructive parameters of the objects. The structure of the parameters allows the most synthetic characterisation of the objects: transport capacity and the cost of technological operations. To solve some planning problems, other characteristics might be necessary as well. Thus, in time the structure of the parameters changes. This means that the form of presentation of the information at the third level has to allow for changes.

4 Reliability of the infrastructure networks

As part of the transport infrastructure, the transport network, similar to other technical infrastructures of the society, can be formally represented by a nonplanar graph with nodes and arcs.

The calculation of reliability between two nodes of the network (connectivity property) is very complex for large networks due to the multiple links (Dupuy [1]). Various mathematical and heuristic models have been developed, and also computer simulation (Bell and Iida [2]). The commonality of all these methods is the fact that in order to establish the probability of having a link between two nodes a structure function is used

$$\Phi(X) = \begin{cases} 1, & \text{if nodes are linked} \\ 0, & \text{if not} \end{cases}$$

where $X = (x_1, x_2, \dots, x_k)$ is a vector with its components representing the operational capacity of different arcs (considering the nodes are ideally operating or that their failures are transferred to the incidental arcs) which represent the link between the points considered, respectively

$$x_i = \begin{cases} 1, & \text{if the arc is operational,} \\ 0, & \text{if the arc is not operational.} \end{cases}$$



Bell and Iida [2] demonstrate that any network can be formally represented with the aid of paths and minimum sections. By path we understand the succession of arcs which ensures the link between the initial and final point, and by section the combination of arcs which when removed from the network causes the interruption of the link.

The minimum path is that path in which the failure of any of the arcs leads to the interruption of the link.

Each minimum path, A_j can be related to the function $\alpha_j(X) = \bigcap_{i \in A_j} x_i$, which takes the value 1 if all the arcs of the minimum path are operating normally, in other words if for all $i \in A_j$ the condition $x_i = 1$ is simultaneously achieved.

The minimum section is that section for which the rehabilitation of the operational capacity of only one arc leads to the rehabilitation of the operational capacity of the whole network. Each minimum section B_k can be related to the logical function $\beta_k(X) = \bigcup_{i \in B_k} x_i$, which takes the 0 value if all the elements belonging to the section fail, and takes the 1 value if at least one element is operating normally.

In general, the network can have several paths and minimum sections. Any network can be represented as a parallel connection of minimum paths or a serial connection of minimum sections. It could be that in a network one and the same element appears in two or more paths or minimum sections. In other words, some minimum paths or/and sections are dependent. This means that network reliability cannot be calculated based on a mathematical model built on the independency of events hypothesis.

It is however possible to evaluate the maximum and minimum limits of the probability of operation without rejection for any size network, with

$$1 - \prod_{j=1}^Q \left(1 - \prod_{i \in A_j} P_i \right) \geq P[\Phi(X)=1] \geq \prod_{k=1}^S \left[1 - \prod_{i \in B_k} (1 - P_i) \right] \quad (1)$$

where P_i is the probability of operation without rejection of element i (e.g., in the transport network case, the number of daily hours without congestion divided by the daily hours the arc of the network is open to the traffic),

Q – number of minimum paths,

S – number of minimum sections.

It can be demonstrated Raicu et al. [3] that in order to simplify the calculations with expression (1) is possible to eliminate the minimum paths and sections with a great number of elements, without substantially altering the precision of the result.

The methods used to evaluate network reliability (irrespective of the nature of the flows) are different from the methods used to calculate the reliability of products. Transport networks, as oppose to telecommunication networks for example have few essential particularities. Some of these particularities are presented hereafter.



5 Particular aspects of transport network reliability

5.1 Reliability in relation to connectivity

The particularity consists in that the effects of congestion on the basis on which the good operation probabilities of the arcs are estimated are also sensed in the nodes of the network affecting the qualitative indicators of transport (time, security, regularity).

5.2 Reliability in relation to capacity

It is defined as being the probability of a transport network to ensure the successful transfer (according to the estimated indicators) of the assigned traffic flows, even when one or several arcs of the network have lost their functionality. It assumes that the reserve of capacity for each arc remains greater than the temporal non-uniformity of the traffic flow or, in a different formulation, or in fact in a different hypothesis – the loss of capacity due to the partial functional degradation of the network cannot be greater than the reserve of capacity.

This characterisation of reliability is necessary to evaluate the transport network. With respect to connectivity the network can ensure the link for any relation between nodes, but this does not mean that the integral transfer of flow is possible in all relations or in a particular one, if one or several arcs have lost their functionality.

A network with good reliability from the point of view of connectivity may not be reliable from the point of view of capacity. In order to characterise the network the concept of vulnerability Taylor and D'Este [4], Berdica [5] or that of operational reliability Raicu [6] has to be employed.

5.3 Reliability in relation to transit time

Coming back to the problem presented before and assuming those reserves of capacity would ensure the operational reliability of the network, it is obvious that the alternative routes suppose greater transit times than the initial ones, maybe even unacceptable for the users.

Therefore, when estimating transit time reliability admissible limits have to be prescribed.

Examining the predicted route as a succession of homogenous sections for which the available statistics provide data about the daily succession of traffic regimes is possible to build the transition diagram of the Markov process associated to the state of the homogenous sections.

In the simplest manner the states of the route encountered by a vehicle when moving in a certain relation can be binary quantified – congested C, respectively not congested, L.

If the probabilities of going from one state to the other are estimated the signal graph can be built (Figure 3) and can evaluate the probability of the route to be in the non congested state after a number of periods, if at the reference moment it was in the same state.



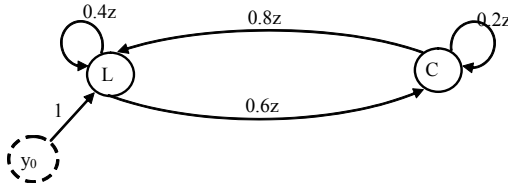


Figure 3: Signal graph associated to a route (the transforms Z of the passing probabilities are written on the arcs).

To solve the problem a dummy node y_0 has been introduced linked to L (“not congested” state) through a dummy arc of transmittance 1.

Mason’s rule is used to evaluate the transmittance. According to this rule, the transmittance between two nodes i and j of the graph is:

$$T_{ij} = \sum \Pi_k \frac{\Delta_k}{D}$$

where Π_k is the transmittance of the direct route k between i and j ,

$$D = 1 - \sum_i L_i + \sum_{i \neq j} L_i \times L_j - \sum_{i \neq j \neq r} L_i \times L_j \times L_r$$

where L_i, L_j, L_r are the transmittances of the loops of the graph, if $L_i \times L_j \times L_r \times \dots$ are achieved only for the non-adjacent loops (those that do not have a common node),

Δ_k has a similar expression as D , except it refers to the loops left after the nodes of the direct route k of Π_k transmittance have been eliminated.

For the graph In Figure 3, $D=1$, $\Pi_{y_0-L}=1$ with $\Delta_1 = 1 - 0.2 \cdot z$. Because

$D = 1 - (0.4z + 0.2z + 0.48z^2) + 0.08z^2$, results

$$T_{y_0-L} = \frac{1 - 0.2z}{1 - 0.6z - 0.4z^2} = 1 + 0.4z + 0.64z^2 + 0.544z^3 + 0.6824z^4 + 0.62704z^5 + 0.649184z^6 + 0.6403264z^7 + 0.64386944z^8 + \dots$$

The coefficients of z^k represent the probability that after a number of k periods the system (analysed route) is in “not congested” state, L . It can be noted that the value of the coefficients of z^k after a number of periods start to stabilise which confirms the fact that state L is a recurring event with a finite average recurrence time.

5.4 Reliability – propriety of the networks

The networks of the technical infrastructures of the society, in general, and the transport networks in particular are topologically and structurally characterised through properties like connexity, connectivity, homogeneity, isotropy, nodality, etc. These properties, as oppose to others like length, density, etc outline the links between the networks and the territorial system they serve [1, 3].



Using a simulation model (written in Visual FoxPro) Raicu et al. [3], the reliability matrix in relation to connectivity is obtained, even for large networks in relatively short computational time.

6 Analysis of rail transport network performance

The specific aim of this research is to develop a suite of decision support tools for assessing network performance and reliability. The tools necessary to analyse and assess existing rail networks, and to plan and evaluate proposed infrastructure improvements and access provision using congestion pricing models will be developed. The outcome will be a set of high-level tools for strategic planning of the long-term development of rail networks.

A particular emphasis is the study of congestion and methods to identify the extent and spread of congestion in a network.

6.1 Congestion/delay in the rail network

The relationship between delay and congestion has been explored. We consider that congestion exists if the passage of one train causes delays to other trains in the system. The level of congestion would then depend on the extent of the delays. Some congestion may be inevitable as a consequence – it is inefficient to design a railway system in which each train has its own, uniquely assigned track. It will be determined if the level of congestion is acceptable or not.

The first observation that should be made is that congestion generates delays, but not all the delays are a consequence of the congestion. Initial delays may arise through some exogenous factor or event, such as a mechanical breakdown. The consequence of these delays may then be the occurrence of congestion.

6.2 Congestion maps

The main sources of train delays are functions of: timetabling; infrastructure or train performance; poor scheduling.

The University of South Australia has developed train scheduling software, internally known as TPOD (Train Plans On Demand) that is able to produce train plans quickly and effectively (see Figure 4). TPOD takes a set of train requests on a particular rail network and minimises the aggregate delays whilst moving the trains from their origins to their destinations Eitzen [7]. By using TPOD to generate the train plans, the poor scheduling factor can be alleviated. If we assume timetabling is largely market driven and therefore largely fixed then delays are an indicator of infrastructure bottlenecks. Delays can be costed (e.g. on an annual basis) yielding a dollar value for capital expenditure justification. Then, the infrastructure can be upgraded to reduce bottlenecks and retest via TPOD to quantify the impact of new infrastructure and confirm cost/benefit (see Figure 5).



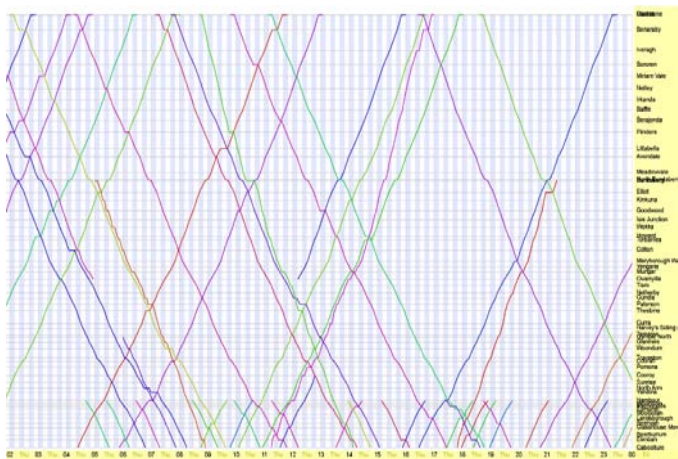


Figure 4: Example of TPOD model output.

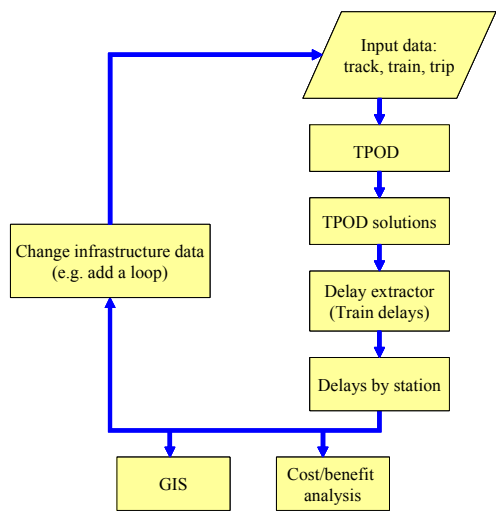


Figure 5: Delay and congestion analysis- system architecture.

A procedure to extract delay information from TPOD (Train Plans on Demand) was designed and programmed. The procedure captures train delays from a TPOD network solution and can display them by station, by time for individual trains or any combination of trains. (Figure 6) The delay data can be exported to a spreadsheet or other planning software for cost/benefit analysis or used to simulate the network working under new conditions for planning purposes and congestion analysis.



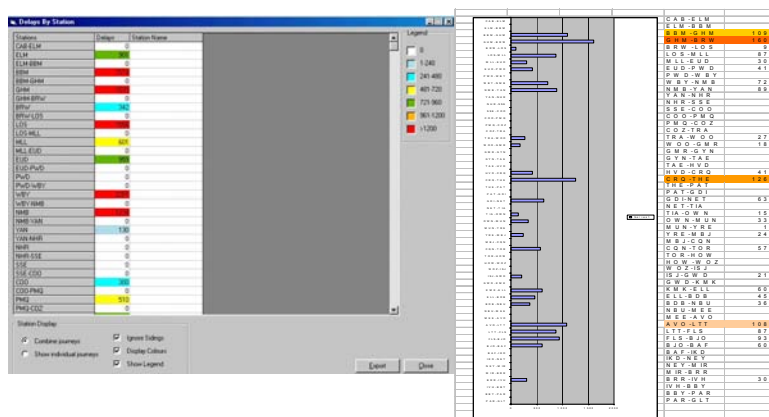


Figure 6: Example of delays by station (output from the delay extractor).

7 Conclusions

This research will lead to the development of expertise in the formulation and application of methods for the analysis of network reliability, identification of vulnerable locations, and assessment of level of performance of intact and degraded networks. This will allow planner and managers to determine the sensitivity of overall network performance to variations in local capacity and local operations, and to identify congestion points and their impact on network performance. These new tools will assist managers and planners to develop insights into network performance under different operating scenarios and objective evaluation of proposed infrastructure development proposals.

References

- [1] Dupuy, G., *Systèmes, réseaux et territoires: principe de reseautique territoriale*. Paris, Presses de l'ENCP, 1985.
- [2] Bell, M.G.H., Iida, Y., *Network Reliability. Transportation Network Analysis*. John Wiley & Sons, pp.179-192, 1997.
- [3] Raicu, R., Raicu, S., and Popa, M., The influence of transportation networks reliability on city logistics. In *Proceedings of the IV International Conference on City Logistics*, 12-14 July, Langkawi, pp 117-131, 2005.
- [4] Taylor, M.A.P., D'Este, G.M., Critical infrastructure and transport network vulnerability. *The Second International Symposium on Transportation Network Reliability (INSTR)*, Christchurch, New Zealand, August, pp.96-103, 2004.
- [5] Berdica, K., An introduction to road vulnerability: what has been done, is done and should be done. *Transport Policy*, **9(2)**, pp.117-127, 2002.
- [6] Raicu, S., Asupra fiabilitatii sistemului de transport feroviar. *Revista Transporturilor*, no. 5-6, pp.145-151, 1991.
- [7] Eitzen, G., TPOD Documentation Draft, University of South Australia, 2004.

A study on a mathematical model of the track maintenance scheduling problem

S. M. Oh^{1,3}, J. H. Lee², B. H. Park¹, H. U. Lee² & S. H. Hong¹

¹*Policy & Operations Research Department,
Korea Railroad Research Institute, Korea*

²*Track & Civil Engineering Research Department,
Korea Railroad Research Institute, Korea*

³*Department of Industrial Systems and Information Engineering,
Korea University, Korea*

Abstract

This paper presents a mathematical model of the long-term track tamping scheduling problem in the Korean high-speed railway system. The presented model encompasses various operational field constraints and moreover improves a state-of-the-art model in extending the feasible space. We show the model is sized up to intractable scale, then propose another approximation model that can be handled with the present computer system and commercial optimization package directly. The aggregated index, lot, is selected, considering the resolution of the planning horizon as well as the scheduling purpose. Lastly, this paper presents two test results for the approximation model. The results show the approximation model to be quite promising for deployment into an operational software program for the long-term track tamping scheduling problem.

Keywords: tamping scheduling, mathematical model, index aggregation.

1 Introduction

In the ballasted track system, the amount of gauge irregularity increases in proportion to the accumulated tonnage passed. When the amount of gauge irregularity is greater than a certain level, it results in not only the worse quality of traveling, but also a serious effect on traveling safety of trains. Therefore, the maintenance works of track systems are carried out based on the predetermined



standards which is set to beforehand prevent from damaging the quality and safety of traveling [1, 2].

Tamping ballast is given the most importance among the maintenance works for correcting gauge irregularity. In Korea Railroad, tamping ballast is performed by hand or by multiple tie tamper (MTT), but MTT is used in most tracks except the some tracks not to be accessible.

Track tamping scheduling problem (TTSP) is to schedule MTT works for maintaining the regular amount of gauge irregularity below the criterion. To do so, target lines are divided by the 'lots' with the previously defined length (the basic length is 0.2 km) and the various parameters of gauge irregularity and actual data are maintained based on the lots. After all, TTSP is to find the schedule of MTT works performed on the lots in order to prevent the standard deviation of gauge irregularity of each lot from being below the criterion.

In this paper, we present mathematical models for TTSP. Many combinatorial problems are intractable if applied to the real field. But our models are developed to be solvable with the current computer system and commercial optimization packages.

This paper is organized as follows; Section 2 presents a mathematical model of TTSP. In Section 3, the mathematical model is scrutinized to devise an approximation model using the index aggregation. Finally, the experimental results and the future research directions are given.

2 Long-term scheduling problem for track tamping

2.1 Outline of the problem

The goal of TTSP is like these; (i) to minimize the number of track tamping works while meeting all requirements related with gauge irregularity (ii) to minimize the fixed costs from every departure for tamping by confirming continuity of each lot (iii) to disperse works not to be concentrated on (limited) specific days. In this section, mathematical model of TTSP is presented. This paper assumes followings for the model.

2.1.1 Assumptions

- i . The proposed model is applied as a unit of railway maintenance post.
- ii . The proposed model uses the concept of 'virtual depot' to take into account multiple maintenance operations during unit period, thus the multiple number of virtual depots may be assigned, if necessary (e.g. 3-5 virtual depots).
- iii . One and only one maintenance operation and MTT type are assigned to each virtual depot during unit period.
- iv . The various rules and processes concerning TTSP are compliant to processes of track tamping which are carried out by 'High Speed Railroad Railway Control Regulation' of Korean Railroad.

The notations used in the model are defined as follows.



2.1.2 Notations

J	The set of virtual depots. The index for indicating the number of maintenance operation(s) in a unit period
L	The set of lots. $L = \{0, 1, 2, \dots, L^{\max}\}$. $L1 = L - \{0\}$. $L2 = L - \{L^{\max}\}$. In here, $L = 0$ indicates depot itself. The indices are sorted in ascending order to outbound direction.
K	Types of MTT equipments. $K = \{\text{new type, old type}\}$.
T	$T = \{1, 2, \dots, T^{\max}\}$. $T^{\max} = 5 \text{ years} \times 365 \text{ days}$.
$\sigma^{\max}$	The standard value of gauge irregularity (standard deviation)
σ_i^+	The increment of gauge irregularity of lot-1 (standard deviation)
σ_k^-	The recovering capacity of MTT type-k of train (standard deviation)
D	Distance matrix for between lot-pair (km)
NW	Prohibition of assigning works.
WT^k	The required operation time of MTT type-k (h/lot)
VR^k	The moving speed of MTT type-k (km/h)
BT	Blocking time (h/day) = working hour + moving time/hour.
UL	The length (km) of a unit lot. Basic length = 0.2 km.
F^k	The fixed costs occurred assignment of MTT with type-k
δ	A very small constant value

2.1.3 Decision variables

$z^{ik} = 0, 1, 2, \dots$	The number of maintenance operations with MTT type-k in period-t.
$I^{jk} = \{0, 1\}$	An indicator function. If a maintenance work is assigned to virtual depot-j during the period-t with a MTT type-k, then it has a value 1, otherwise 0.
$w_l^{jkt} \in \{0, 1\}$	If a maintenance work is assigned to lot-l by a virtual depot-j using MTT type-k during the period-t, then it has a value 1, otherwise 0.
$y_{lm}^{jkt} \in \{0, 1\}$	A variable for a sequence of maintenance works assignments. If $w_l^{jkt} = 1 \Rightarrow w_m^{jkt} = 1$, then it has a value 1, otherwise 0.
$N^{jkt} \equiv \{l \mid w_l^{jkt} = 1, \forall j \in J, \forall k \in K, \forall t \in T\}$	A set of lots where a maintenance work is assigned for virtual depot-j with MTT type-k in period-t, dynamically. $n = N^{jkt} $
$\mu^{jkt} = \max \{l \mid l \in N^{jkt}\}$	
$\nu^{jkt} = \min \{l \mid l \in N^{jkt}\}$	
$\sigma_l^t \in R^+$	The levels of gauge irregularity in lot-l at the end of period-t (standard deviation)
$Z_1 \in R^+$	The first term of objective function in TTSP0 (variable cost)
$Z_2 \in R^+$	The second term of objective function in TTSP0 (fixed cost)



2.2 Objective function

The objective function of TTSP model consists of two terms. The first one is to minimize the frequency of maintenance operations while meeting all requirements. In the case that the amount of gauge irregularity exceeds the criterion, the maintenance operations are indispensable but the number of the operations should be minimized because each operation makes the ballast crushed. Formula (1) is the function to achieve such objective.

$$Z_1 = \sum_{i \in T} \sum_{k \in K} \sum_{j \in J} \sum_{l \in L} \sum_{m \in L_1} (w_m^{jkt} - \delta \cdot y_{lm}^{jkt}) \quad (1)$$

The second term is to dispose the given works with the minimum number of maintenance works. It is not common that more than maintenance work group is operated in a day at one depot. Therefore, the number of maintenance work groups might be minimized instead of minimizing the number of departure of maintenance works.

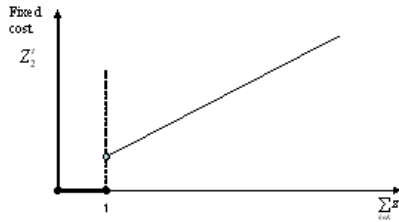


Figure 1: An example of fixed cost.

$$Z'_2 = \begin{cases} 0 & \sum_{k \in K} z^{tk} \leq 1 \\ F^k \sum_{k \in K} z^{tk} & \sum_{k \in K} z^{tk} > 1 \end{cases} \quad (2)$$

$$Z = Z_1 + \sum_{i \in T} Z'_2 \quad (3)$$

These objectives can be achieved, as shown in Figure 1, by considering the fixed cost paid whenever maintenance departures more than two are employed during unit period. Thus, the objective function is given, as in formula (3), by the sum of variable cost and fixed cost. To implement the fixed cost in a commercial optimization software package, it needs to modify model. The modification of model will be described in next section.

2.3 Assignments of maintenance operations

It is crucial in TTSP model to keep the amount of gauge irregularity of each lot below the criterion value, which is expressed as the inequality (4).



$$\sigma'_i \leq \sigma^{\max}, \quad \forall t \in T, \forall l \in L \quad (4)$$

The maintenance works are avoided being assigned in uniform fashion, since the increasing rates of gauge irregularity (σ_i^+) are considerably different even in the adjacent lots [3]. It is desirable to establish works which let the periodicity of consecutive lots synchronized with similar periods.

Gauge irregularity of each lot increases at every unit period like formula (5-6), and track tamping works (w_i^{jkt}) will be established when the values are reached at regulated level ($\sigma^{\max}$). Meanwhile, formula (6) restricts the feasible space to be limited. Figure 2 depicts the restriction conceptually.

$$\sigma_i^1 = \sigma^{\text{init}}, \quad \forall l \in L \quad (5)$$

$$\sigma_i^t = \sigma_i^{t-1} + \sigma_i^+ - \sum_{k \in K} \sum_{j \in J} \sigma_k^- \cdot w_i^{jkt}, \quad \forall t = \{2, 3, \dots, T^{\max}\}, \forall l \in L \quad (6)$$

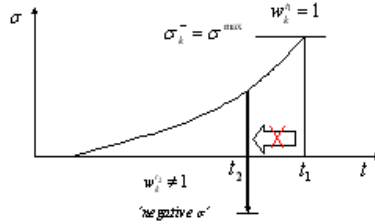


Figure 2: Restriction of feasible space.

When assigning maintenance operations with assumption that corrective effects of gauge irregularity is set exactly same as the regulated level (i.e. $\sigma_k^- = \sigma^{\max}$), assigning operations is possible ($w_i^1 = 1$) only when the amount of gauge irregularity reaches exactly at the standard level ($t = t_1$), and all previous operation assignments are impossible ($w_i^2 \neq 1$). This is because operations are assigned in advance of t_1 , $\sigma_i^t \geq 0$ is violated. This paper presents a modified formulation for this point in formula (7).

$$\sigma_i^t = \sigma_i^{t-1} + \sigma_i^+ - (\sigma_i^{t-1} + \sigma_i^+) \cdot \sum_{j \in J} \sum_{k \in K} w_i^{jkt}, \quad \forall t = \{2, 3, \dots, T^{\max}\}, \forall l \in L \quad (7)$$

On contrary to formula (6), if necessary, formula (7) allows operation assignments possible even before the date of reaching to standard level of gauge irregularity. It means that a limitation which was discovered in the state-of-the-art study [1] becomes relaxed. To apply commercial optimization package, further modification of the formula is necessary because decision variables in the right side take the form of multiplication. The modification of model will be described in the next section as that of the previous one.



$$\sum_{k \in K} \sum_{j \in J} w_l^{jkt} \leq 1, \quad \forall t \in T, \forall l \in L \quad (8)$$

$$\sum_{k \in K} \sum_{j \in J} w_l^{jkt} = 0, \quad \forall (t, l) \in NW \quad (9)$$

Formula (8) ensures the uniqueness of operations which are assigned in every unit period and lot, and formula (9) is to restrict assignment of operations owing to the operational reasons during cold and hot seasons.

2.4 Restrictions on blocking times

For all lots, the sum of working time and moving time is restricted by blocking time (BT). Figure 3a represents an example that a team of maintenance operation departs the depot, processes/moves lots, the turns back. Such working process can be expressed by a graph of which assignment of operation is node, and movements between assignments are arcs as shown in Figure 3b.

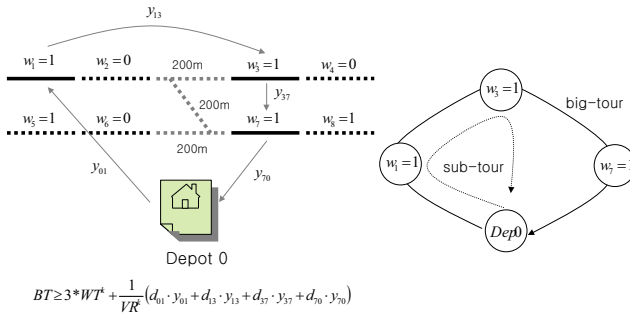


Figure 3: (a) An example of maintenance operation sequence, and (b) its graph representation.

That graph comes under the traveling salesman problem (TSP) which is widely known in the field of combinational optimization. The blocking time of the graph operates only for the big-tour that visits all nodes with a shortest path.

$$\sum_{l \in L_2} \sum_{m \in L_1} \left[(y_{lm}^{jkt} - 1) WT^k + \frac{d_{lm} y_{lm}^{jkt}}{VR^k} \right] \leq BT, \quad \forall j \in J, \forall t \in T, \forall k \in K \quad (10)$$

$$\sum_{l \in S^{kt}} \sum_{m \in S^{kt}} y_{lm}^{jkt} \leq |S^{kt}| - 1, \quad \forall S^{kt} \subset N^{jkt}, 2 \leq |S^{kt}| \leq n - 1 \quad (11)$$

$$\mu^{jkt} \leq \nu^{jkt} + 1, \quad \forall j \in J, \forall k \in K, \forall t \in T \quad (12)$$

$$w_l^{jkt} + w_m^{jkt} - 2y_{lm}^{jkt} \geq 0, \quad \forall j \in J, \forall k \in K, \forall t \in T, \forall l \in L_2, \forall m \in L_1 \quad (13)$$

Formula (10) represents the constraints of blocking time on big-tours, and formula (11) is sub-tour elimination constraints. These constraints are known as make models very difficult (NP-Complete) to be solved [4]. Formula (12) is a constraint to avoid duplication of operation-assignment-range in each virtual



depot. Formula (13) defines the logical connection between operation assignment variables ($\mathbf{w}$) and operation sequence variables ($\mathbf{y}$), and it works in proper for the minimization form of objective functions.

2.5 Calculating frequency of operations

In the objective functions described in previous sections, to optimize the number of maintenance departures during a unit period, frequency of operations should be calculated.

$$I^{jkt} - w_l^{jkt} \geq 0, \forall t \in T, \forall j \in J, \forall k \in K, \forall l \in L \quad (14)$$

$$\sum_{k \in K} I^{jkt} \leq 1, \forall t \in T, \forall j \in J \quad (15)$$

$$z^{tk} - \sum_{j \in J} I^{jkt} = 0, \forall t \in T, \forall k \in K \quad (16)$$

Formulas (14-16) is indicator function which is established in accordance with operations in specific lots, and the operation frequency of each MTT type during unit period is calculated by formula (16).

2.6 The range of variables

Formulas (17-21) represents the range of decision variables.

$$z^{tk} = 0, 1, 2, \dots \quad (17)$$

$$I^{jkt} \in \{0, 1\} \quad (18)$$

$$w_l^{jkt} \in \{0, 1\} \quad (19)$$

$$y_{lm}^{jkt} \in \{0, 1\} \quad (20)$$

$$\sigma_i^t \in R^+ \quad (21)$$

Long-term scheduling model of ballasted track tamping is defined from those defined objective function(s) and constraints as follows.

(TTSP0) - Track Tamping Scheduling Problem
Minimize (3)
Subject to (4-5), (7-21)

3 Approximation of model

3.1 Aggregation of index

Problem size of TTSP0 whose subjects are the newly constructed Korean high speed line (HSL) between Busan and Seoul can be estimated based on the most difficult variables (y_{lm}^{jkt}).



$$\begin{aligned}
y_{lm}^{/kt} &\Rightarrow [3-5] \text{ (departures of maintenance operations)} \times 1825 \text{ (periods)} \\
&\times 2 \text{ (MTT types)} \times 1105 \text{ (lots)} \times 1105 \\
&= 22,283,706,250 \Rightarrow 2^{22,283,706,250}
\end{aligned}$$

With the computability of computer systems currently used, this scale of problem is not easy to solve directly. Index aggregation is one of the potential approaches to these large-scaled models. The index aggregation means to reduce scales of the models that can be manipulated using current computer systems, by cutting the number of variables within the range of keeping actual meaning of solutions as far as possible. It seems excessively specific to keep the decision values of long-term scheduling problems for several decades of kilometers by the unit of 0.2 km. In this paper presents a solution approach that integrates those excessively specific decision variables.

As for the operational speeds of MTT equipments which are currently used by Korean Railroad, those of new equipments are about 1.0 km/h (= 5 lots/h, 0.2 km standard), those of old equipments are about 0.8 km/h (= 4 lots/h, 0.2 km standard). According to those operational, on the assumption that average blocking time a day is 4 h (the average processing time a day = the average blocking time a day - the average running time a day = 4 h - 1.5 h = 2.5 h), the average processing lots of new equipments a day is about 12.5 lots/day (= 5 lots $\times$ 2.5 h/day), the average processing lots of old equipments is about 10 lots/day (= 4 lots $\times$ 2.5 h/day). The number of average lots of new and old equipments is about 11 lots a day. Therefore, in this paper, we propose Aggregated LOT (AL) which integrates lots of 0.2 km (11 lots) which can be operated for a day.

By integrating index of lots, the types of the equipments and the concept of virtual depot of TTSP0 is not necessary any more, and decision parameter associated to lots ($l \in L$) may be re-defined as average value (s_a^t) for aggregated lots ($a \in AL$). By aggregating index of lots, the number of variables can be reduced considerably, and relevant formulas (10-13) can be relaxed as well. In the next section, Approximated TTSP (TTSPA) based on the index aggregation is presented.

3.2 Objective function of TTSPA

In the approximated model, the objective function is modified using newly defined variables. First, objective (2) concerning variable cost of TTSP0 is defined as formula (22) in TTSPA.

$$Z_3 = \sum_{t \in T} \sum_{a \in AL} c^t \cdot w_a^t \quad (22)$$

The cost weight (c^t) concerning operation assignment of formula (22) is to solve problem arising from making problem of infinite time horizon smaller to problem of finite time horizon. For example, in the problems of definite time horizon as Figure 4a, operation assignments are possible regardless of the



degree(s) of access to the standard (value) at a middle spot, within the range of meeting the standard amount of gauge irregularity in scheduling areas. In the case of infinite time horizon, assigned operations in a cycle are logically right to assign near the standard approximation except for the invincible situations.

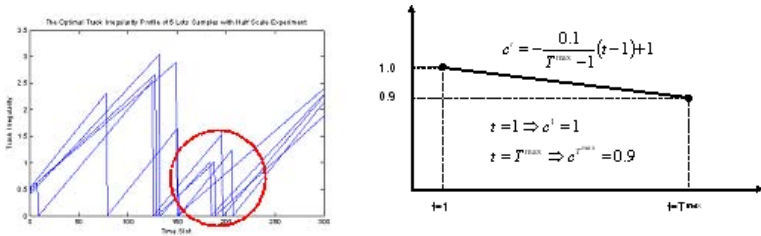


Figure 4: (a) Assignments in infinite time horizon, and (b) Weights to compensate.

The weight concerning assignment of operations compensates such problems arising from the models of finite time horizon, inducing assignment of in the latter half of the period, approaching to the standard. Such weight should be defined in the range not to influence on the original objective function defined by users. Figure 4b depicts an example of methods for establishing weights.

Modification of Formula (2), objective function concerning fixed costs of TTSP0, is for manipulating in commercial optimization software packages. The objective functions which have fixed costs are able to be realized by using Piece-wise Linear (PWL). The PWL is an efficient method which helps realize fixed costs with a small increment number of variables [5].

$$z'_{41} + z'_{42} - \sum_{a \in AL} w_a^t = 0, \forall t \in T \quad (23)$$

$$0 \leq z'_{41} \leq 1, \forall t \in T \quad (24)$$

$$0 \leq z'_{42} \leq z^{\max} - 1, \forall t \in T \quad (25)$$

$$Z_4 = F_{41} \sum_{t \in T} z'_{41} + F_{42} \sum_{t \in T} z'_{42} \quad (26)$$

$$Z = Z_3 + Z_4 \quad (27)$$

The objective function is defined as formula (27) by formulas (22) and (26).

3.3 Linearization assignment constraints

Although formula (7) is an alternative to the problem of restrictive feasible space, some problems in the third term of right-hand side arise when two decision variables are expressed as the forms of multiplication, and apply a



commercial optimization software packages to the problem. Such a quadratic model can be linearized as follows. First, formula (28) is given by using the variables defined by aggregated index.

$$s_a^t = s_a^{t-1} + s_a^+ - (s_a^{t-1} + s_a^+) \cdot w_a^t, \forall t = 2, \dots, T^{\max}, \forall a \in AL \quad (28)$$

When $\Delta_a^t = s_a^{t-1} \cdot w_a^t$, formula (28) is linearized as formulas (29-33).

$$s_a^t = s_a^{t-1} + s_a^+ - \Delta_a^t - s_a^+ \cdot w_a^t, \forall t = 2, \dots, T^{\max}, \forall a \in AL \quad (29)$$

$$\Delta_a^t \geq 0, \quad \forall t = 2, \dots, T^{\max}, \forall a \in AL \quad (30)$$

$$\Delta_a^t \leq s^- \cdot w_a^t, \quad \forall t = 2, \dots, T^{\max}, \forall a \in AL \quad (31)$$

$$\Delta_a^t \leq s_a^{t-1} + (1 - w_a^t) \cdot s^-, \quad \forall t = 2, \dots, T^{\max}, \forall a \in AL \quad (32)$$

$$\Delta_a^t \geq s_a^{t-1} - (1 - w_a^t) \cdot s^-, \quad \forall t = 2, \dots, T^{\max}, a \in AL \quad (33)$$

To make understand clear about the model, some formulas are repeated, and TTSPA model is as follows.

(TTSPA)

$$\text{Minimize } Z = Z_3 + Z_4 \quad (27)$$

Subject to

$$Z_3 = \sum_{t \in T} \sum_{a \in AL} c^t \cdot w_a^t \quad (22)$$

$$z_{41}^t + z_{42}^t - \sum_{a \in AL} w_a^t = 0, \quad \forall t = T \quad (23)$$

$$0 \leq z_{41}^t \leq 1, \quad \forall t \in T \quad (24)$$

$$0 \leq z_{42}^t \leq z^{\max} - 1, \quad \forall t \in T \quad (25)$$

$$Z_4 = F_{41} \sum_{t \in T} z_{41}^t + F_{42} \sum_{t \in T} z_{42}^t \quad (26)$$

$$s_a^t = s_a^{t-1} + s_a^+ - \Delta_a^t - s_a^+ \cdot w_a^t, \quad \forall t = 2, \dots, T^{\max}, \forall a \in AL \quad (29)$$

$$\Delta_a^t \geq 0, \quad \forall t = 2, \dots, T^{\max}, \forall a \in AL \quad (30)$$

$$\Delta_a^t \leq s^- \cdot w_a^t, \quad \forall t = 2, \dots, T^{\max}, \forall a \in AL \quad (31)$$

$$\Delta_a^t \leq s_a^{t-1} + (1 - w_a^t) \cdot s^-, \quad \forall t = 2, \dots, T^{\max}, \forall a \in AL \quad (32)$$

$$\Delta_a^t \geq s_a^{t-1} - (1 - w_a^t) \cdot s^-, \quad \forall t = 2, \dots, T^{\max}, a \in AL \quad (33)$$

$$w_a^t = 0, \quad \forall (t, a) \in A^{\text{WN}} \quad (34)$$

$$z^t = 0, 1, 2, \dots, z^{\max} \quad (35)$$

$$w_a^t \in \{0, 1\}, \quad \forall t \in T, \forall a \in AL \quad (36)$$

$$0 \leq s_a^t \leq s^{\max} \quad (37)$$



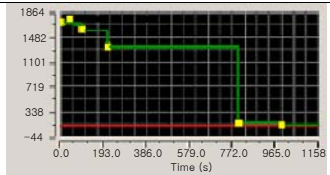
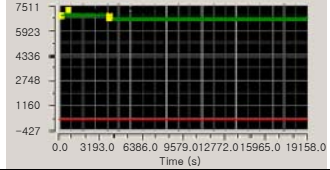
In this section, an approximated model is presented to be handled by using current computers and commercial optimization software packages. In the next section, results of numerical experiments and hereafter direction concerning TTSPA are presented.

4 Experimental results and conclusions

4.1 Experimental results

Two experiments on TTSPA models were performed as shown in Table 1. The one is an experiment about average track tamping problem size (100%) of Korean HSL at each maintenance depot, and the other is a 60%-scaled model of the problem size. They were performed twice for each, and the result of each performance was similar. The result of Table 1 was got from using OPL Studio 4.0 (CPLEX 9.0 is built in) of ILOG Co. in a PC which has 3.4 GHz CPU, 1Gb RAM. Table 1 shows the summary of experiments.

Table 1: Experimental results of TTSPA.

Problem size	Solutions	Experimental conditions
60%		- CPU time = 772 s. - GAP = 15% - Trials = 2 times - Constraints = 60,250 - Int. variables = 45,600
100%		- CPU time = 6 h - GAP = 96.5% - Trials = 2 times - Constraints = 202,399 - Int. variables = 152,500

4.2 Conclusions

TTSPA came to within a scale of commercial optimization software package. Even though the proposed TTSPA model comes into the range where model is manipulated by using current computer systems, further studies are required to improve the convenience of an application program using this model. For the improvement of algorithm performance, further studies for user cuts to improve lower bounds of CPLEX Branch & Cut, and for developing heuristic algorithm to improve effective upper bounds.

On the aspect of modeling, some additional studies about effective reduction by using restrictions of assignment of hot and cold seasons, and about extension to the track tamping model related to ballast cleaning are required.



References

- [1] Miwa, M, Mathematical programming model analysis for the optimal track maintenance schedule, 2002, QR of RTRI, 43(3), pp. 131-136.
- [2] Jae-Hee Kim, Seog-Moon Oh, Jeeha Lee, Soon-Heum Hong, Hee-Up Lee, A Review on Optimal Track Maintenance Scheduling Problem, Proceeding of Korean Society of Transportation, 2004.
- [3] Ji-Tek Oh, A Study on the Planning for Mechanization and Modernization of Track Maintenance, internal report of Korea Railroad Research Institute, 2000.
- [4] Wolsey, L. A., Integer Programming, Wiley Inter-Science, New York, 1998.
- [5] Williams, H. A., Model Building in Mathematical Programming 4th, John Wiley & Sons, New York, 2005.



Statistical method for the evaluation of railway systems modifications

M. Chandesris

*SNCF Direction de la Recherche et de la Technologie,
Unité de Recherche “Génie Décisionnel Appliqué”, Paris, France*

Abstract

Railway production plans daily face random events. Therefore, before setting up modifications for the railway system (changing the network, new signalling system, new timetable, etc.), its robustness to stochastic events must be estimated.

Tests cannot actually be carried out, which is why the SNCF uses a traffic simulator developed by its research department. Up to now, the timetable robustness has been estimated by considering a few small incidents and their simulated consequences on train delays.

The new approach we are working on consists of taking into account the stochastic dimension in the evaluation of robustness. The methodology is divided into three parts:

- Identifying randomness to be taken into account when evaluating robustness
- Studying the robustness of the railway system:
 - determine experimental design
 - carry out simulations
 - aggregate results
- Evaluating the lack of precision in the results

The aim of this project is to develop the required statistical methods and to design a decision tool following the three steps above. It will help the infrastructure managers to evaluate, in a statistical manner, the robustness of various railway system configurations.

1 Timetable robustness and traffic simulator

Railway production plans daily face random events. Therefore, before setting up modifications for the railway system (changing the network, new signalisation



system, new timetable, etc.), its robustness to stochastic events must be estimated.

This paper is divided into three parts:

- Introduction to timetable robustness and the use of traffic simulator
- Proposition for methodology to evaluate robustness of timetable
- Examples of comparing different alternatives for a railway system thanks to the tool.

This first section introduces timetable robustness, simulation model and stochastic dimension of the problem.

1.1 Timetable robustness

Railway production daily faces random events that perturb traffic quality. When evaluating new production plans, these random events have to be included in the analysis of quality.

Such an approach allows introducing more elements when evaluating different modifications for the railway system (increase in traffic density, modification of signalling systems, etc.). Costs, theoretical capacity indicators may be available a priori, directly computable by looking at proposals. But indicators of service quality such as regularity involve randomness phenomena, which is more difficult to take into account in computations.

Timetable robustness is the capacity of a timetable to resist to little perturbations. (We concentrate on little perturbation, considering situations without active regulation.)

1.2 Simulation model

Actual tests cannot be carried out due to several reasons: security, costs, time, etc. Furthermore, the railway system is too complex to be solved by analytical solution. Simulation techniques provide a solution to these problems. The SNCF research department has developed a numerical simulator of the railway system [3]. The simulation model is briefly described.

The railway system such as it is described in SISYFE (SImulateur de SYstème FERroviaire), is comprised of two main subsystems that interact:

- the rolling stock with its driver and the fixed installations of the rail network;
- the human action of men outside this system (signalmen or traffic managers) classified as its environment.

The simulation allows, by iterative calculations, to simultaneously reproduce the changing state of signals and the evolution of the vehicles. The railway system is modelled by a discrete events system. At every moment, the state of the system is determined by:

- former states;
- actions of the environment;

It is a quasi-deterministic system because relation of transition checks consistency conditions. Propagation of state changes generates conditions that will characterise the following states.



Such a tool allows studies of the timetable with several hundred of trains during hours.

The choice of a simulation tool is not restrictive; other simulation tools could be used instead of SISYFE. A lot of simulation tools have been recently developed by or for railway companies. But, nowadays, the key question is no longer “how to build a traffic simulator and how to choose a simulation model?” but “which scenarios have to be simulated into the simulator?”.

1.3 Stochastic dimension

The natural randomness of the production process has to be taken into account. The Monte Carlo methods [4] provide approximate solutions to such problems by performing statistical sampling experiments on a computer.

The idea is to sample a perturbed timetable in an efficient way, then put them into the simulation tool, and then analyse the way perturbations are absorbed and finally aggregate results to make statistical conclusions.

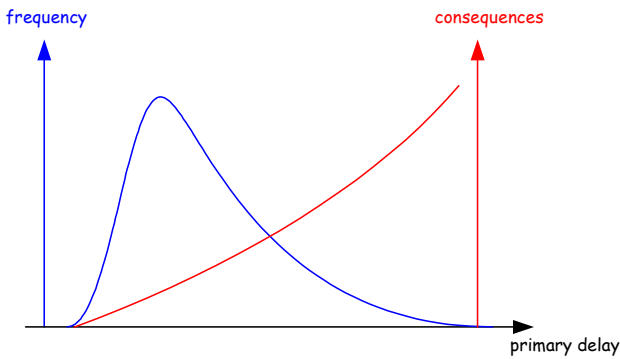


Figure 1: Duration of primary delay occurrence frequency and operational consequences.

2 Evaluation of timetable robustness – methodology

This second section describes how to:

- identify randomness;
- study robustness by simulation;
- evaluate a lack of precision in the results.

2.1 Identify randomness

The first task is to identify random events that would be taken into account during the analysis. Two points of view should be considered:

- incidents that occur during actual exploitation of the system are listed in databases from the feedback. Those databases allow the classification of incidents by type and to study their occurrence frequency;



- incidents that do not cause any perturbation to the system are not to be included in the analysis, in order to optimise the use of simulation. A first set of simulated incidents makes it possible to identify characteristics of the incidents that cause operational consequences.

By studying these two aspects of random events that occur during railway production, we can –in an optimal way– determine which incidents should be taken into account during evaluation of robustness.

2.2 Study robustness of the railway system

2.2.1 Determine experimental design

Simulations have to be carefully planned, in order to best use the simulation time. The theory of experimental design [1] contains very useful techniques. The full design is divided into two parts:

1. the main design produces a *vary hypothesis for the railway system* (configuration parameters). At this step, we must set what we want to test: new timetables with increasing traffic density, new signalling system, new infrastructures, etc.
2. the second design *varies the characteristics of incidents*. Several characteristics of incident may be taking into account:
 - type of the incident;
 - delayed departure;
 - speed limitation for a train or a portion;
 - unexpected stop;
 - emergency stop;
 - primary delay;
 - place where the incident occurs;
 - profile of the line where the incidents occurs;
 - speed of the train when the incident occurs;
 - space margin of the train when the incident occurs;
 - rolling stock of the first train affected;

The building of the incidents design is helped by the former step “identify randomness”.

Each configuration will be tested on the same types of incidents (but not necessarily the same incidents). When joining the experimental designs for configuration and for incidents, a cross-over design is obtained:

Establishing the experimental design is a key step: at this moment we optimise the compromise between:

- number of simulations;
- representative of incidents;
- accuracy of the results.



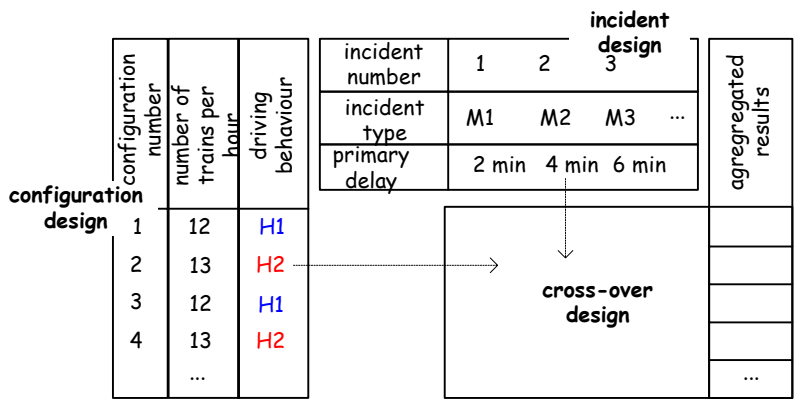


Figure 2: Example of experimental design.

2.2.2 Carry out simulations

SISYFE is based on a very detailed model of the railway system: numerous different configuration hypothesis and types of incidents may be tested. It provides hypothesis trees in order to automatically launch numerous sets of simulations with various hypotheses. We have been working on parallel auto-repartition in order to save time when simulating hundred even thousand simulations.

For each pair (configuration, incident) simulated, SISYFE provides the transport plan “as if” it occurred in such conditions. The life for each train is available: where it is along its path and at what time. Secondary delays are deduced by comparison with theoretical schedule. A lot of numerical and graphics indicators are proposed to analyse this result.

2.2.3 Aggregate results and analysis

The next step is to aggregate results. Two ideas should be kept in mind while the simulation results are aggregated:

- all incidents have to be aggregated by configuration in order to compare configurations between each other.
- incidents must be weighted (by frequency) in order to obtain actual quality indicator values;

To test significance of parameter variations such as traffic density, statistical techniques are used. The main one is variance analysis; the main concepts are explained below.

2.2.3.1 Variance analysis Take the example of the comparison of two configurations for the system. The first configuration is S0, generally the actual railway system configuration. By simulating incidents with this configuration, the distribution of the quality indicator studied is obtained. We write it $S0(I) \rightarrow R + \epsilon$. With 90% chance, the quality indicator is between $R - \epsilon$ and $R + \epsilon$.

When testing an improvement of the actual configuration, we test $S1 = S0 + \Delta S$. The new quality indicator is R1. $[S0 + \Delta S](I) \rightarrow R1 + \epsilon = R + \Delta R + \epsilon$.

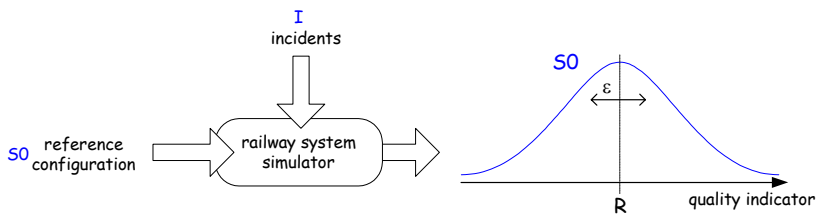


Figure 3: Distribution of the quality indicator obtained by simulations.

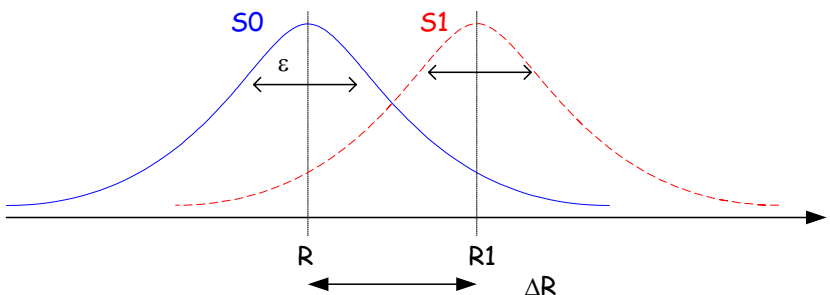


Figure 4: Comparison between two distributions of quality indicator.

The objective is to identify if the modification of the railway system is significant, whether is it worth implementing the modification? The two situations are judged by comparing the variability term and the quality improvement ΔR . The modification is significant if $\Delta R > \epsilon$.

2.3 Evaluating lack of precision in the results

Another crucial point is the estimation of the error of the present method. When announcing a rate of regularity under hypothesis of configuration, a confidence interval is added. It is more pragmatic and more interesting to provide piece of information of type: the rate of regularity is between R1 and R2 than to provide only R.

This is illustrated by the following example: we wish to increase the number of trains per hour. But, if the number of trains per hour is increased, delays are spread more easily because trains are closer together. So we want to test if this new delay can be compensated for by changing the trains are driven. Two hypotheses, H1 and H2, have to be tested.

At first glance at Figure 5, if the actual situation is 12 trains per hour with H2 (red in dot line), the number of trains per hour can be increased without losing regularity if the driving behaviour changes to H1 (blue in plain line).

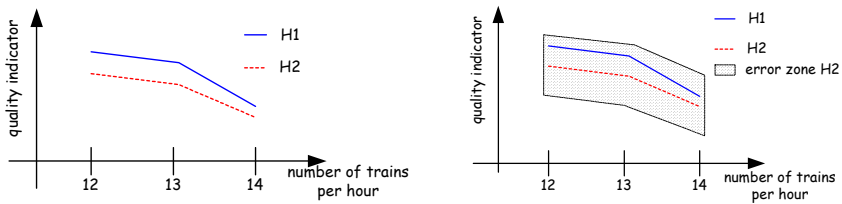


Figure 5: Interest of providing confidence interval.

But, considering the confidence interval of the evaluation process, this special case cannot be concluded, because the error made on the evaluation is greater than the difference between the two hypotheses.

If different hypotheses have to be compared in order to take investment decision, the gap between indicators of each hypothesis have to be bigger than the precision of the process. Otherwise, we cannot be sure of results of the decision.

The lack of precision in the process of measurement of robustness comes from three main sources:

1. in the feedback databases, the estimation of occurrence frequency of incident types may be biased;
2. a simulation model is a simplification of the actual world;
3. only a finite number of incidents are simulated.

Finally, the lack of precision may be evaluated by:

- standard parametric techniques such as confidence interval if normality assumptions are fulfilled;
- non parametric techniques, bootstrap may be used for example [2].

We have been working on both methods to obtain the smallest error term.

3 Compare different alternatives for a railway system

Such a methodology provides a decision indicator for service quality. In this section, we expose some examples of improvements of railway system that could be evaluated through this procedure. Then the tool that has been developed is briefly presented.

3.1 Examples of studies

3.1.1 Adding infrastructure

Such tool may be used to evaluate the contribution of a new infrastructure. For example, the regularity with and without a third track could be compared. Then the quality contribution of this new track could be compared to the economic costs.



3.1.2 Increasing traffic density

A first study on the high-speed line between Paris and Lyon has been done, taking into account about 5 000 simulations. The influence of two parameters have been tested:

- the number of trains per hour : 12, 13, 14 or 15 trains/hour;
- two driving behaviour parameters: A1, A2 and L1, L2.

3.1.3 Modifying signalling and / or routing systems

Another study involves an interconnection project in Paris and suburbs. Hundreds of trains of different speed are involved. We want to evaluate the contribution of the ETCS 2 signalling system with manual or automatic driving. Different systems of regulation have to be tested too.

3.2 A statistical tool to help decision makers

A prototype tool, SARDAIGNE, has been developed in SAS (Statistical Analyses Software, see www.sas.com), including a graphical interface for the user. It is based on the proposed methodology and is divided in four main parts:

- import and data management
- analyse for one simulation
- production of list of scenarios in order to test factors influence
- analyse of list of simulations

4 Conclusions

To conclude, timetable robustness, the capacity of a timetable to resist small perturbations, has to be estimated by simulation before applying modifications to the railway system (increase of traffic density, modification of signalling system, etc.). The methodology has to include a probabilistic/statistical view of perturbations of railway system.

In this paper, we described the main ideas of how to identify randomness, how to study robustness by simulation and how to evaluate a lack of precision in the results.

Some examples cases are given. A tool for providing such decision help has been developed.

References

- [1] D. Benoist, Yves Tourbier and S. Germain-Tourbier, *Plans d'expériences: construction et analyse*, Technique & documentation, Lavoisier, 1994.
- [2] B. Efron, R. J. Tibshirani, *An introduction to the bootstrap*, Chapman & Hall, 1993
- [3] M. Fontaine and D. Gauyacq, *SISYFE: a toolbox to simulate the railway network functioning for many purposes. Some cases of application*, World Congress on Railway Research, 2001
- [4] G.S. Fishman, *Monte Carlo*, Concepts, algorithms and applications, Springer series in operations research, Springer, 1996.



Prioritized Rail Corridor Asset Management

T. Selig¹ & J. Kasper²

¹*Optram Inc., Maynard, Massachusetts, USA*

²*MRO Software, Inc., Bedford, Massachusetts, USA*

Abstract

Seldom do railroads have the resources to maintain their infrastructure at such a level to ensure steady-state performance. More often they are faced with the decision of which maintenance actions should take priority, so that optimal safety and reliability are maintained under the burden of constrained resources. The option of additional resources is rarely feasible. Given these current operating realities, more and more railroads are finding that the solution lies in utilizing information technology to work more efficiently. By harnessing the vast amount of existing rail corridor data into a prioritized plan, and then assigning the work and monitoring the execution and results with software tools, many railroads are attempting to do more with less. The strategy, we define as Prioritized Rail Corridor Asset Management, not only results in steady-state asset performance under constrained resources but can even improve the asset condition and provide a positive return on investment.

This paper describes an information system developed to facilitate a Prioritized Rail Corridor Asset Management strategy. The enterprise-scale solution requires combining Optram's Railway Infrastructure Management (ORIMTM) information system with MRO Software, Inc. (MROI) strategic asset management information system (MAXIMO[®]). ORIM integrates existing rail corridor data sources with a linear model of the track network to continuously generate a prioritized work plan. The prioritized work plan is linked to MAXIMO where staff, materials, and equipment are managed. The results combined with an effective execution facilitate a Prioritized Rail Corridor Asset Management strategy.

Keywords: Railway Infrastructure Asset Management, Enterprise Asset Management, Linear Asset Management, Track Maintenance Management, Rail Management, Enterprise Information Systems.



1 The Challenge

Managing a reliable and safe rail corridor is typically performed with insufficient information and limited resources. Having simple to use and readily available asset location and condition information to prioritize the use of limited resources (people, material, equipment, and work windows) can dramatically affect the rail corridor reliability, safety and profitability. Compounding the problem is a continuous flood of data collected from activities such as track inspections, geometry car surveys, rail defect detectors, ground penetrating radar, rail profile measurements, video surveys, infrared tests, asset surveys, work records, train movement, wayside detectors and more. Converting this vast, continuous and varied stream of data into management intelligence to develop a prioritized plan to optimize maintenance-of-way resources is a significant challenge.

Successful applications of prioritized rail corridor asset management can have significant and lasting impact. It has been shown at Amtrak and Conrail [1] and at Network Rail and First Engineering [2] that without additional capital investment a 10% annual savings in maintenance expense could be saved by targeting timing and location of capital and maintenance operations and resources [1]. Higher returns can be realized when capital investments are also made in the infrastructure.

2 Rail Corridor characteristics

2.1 Rail Corridors as a linear asset

Rail corridors can best be characterized as comprised of a network of interrelated linear or “continuous” assets. The linear asset is similar to other capital-intensive assets—such as facilities and fleets of vehicles—in that they are expensive to build and maintain, and have a direct impact on rail service reliability, safety and profitability. On the other hand, there are important differentiations for linear assets:

- They are geographically dispersed, with few or no obvious physical characteristics that distinguish one asset from another, other than their relative location
- Maintenance work and component wear is *distributed* over varying portions of the asset in disperse and non-discrete locations along the asset;
- Linear assets are comprised of an interrelated series of components (rail, sleepers/ties, fasteners, ballast, subballast) that vary along the length of the asset
 - The performance of a part of one asset may affect part of another (e.g., a wet spot could increase rail defect rate, bridges transitions, switch and crossings, curves)
 - Work is performed on a portion of an asset on various components
 - Use characteristics (speed, tonnage, etc.) vary by asset location and component



In short, the unique continuous and interrelated nature of linear assets requires a specialized approach to develop and manage a safe, cost-effective and reliable maintenance-of-way plan.

2.2 Impact of Rail Corridor Asset Performance

A railway system is made up of vehicles, facilities, and rail corridor assets. A failure in a vehicle asset can disrupt a single carload and possibly a comprising train making a minor impact on a corridor's throughput. Similarly a failure in facility typically has a minor impact on the systems overall performance. On the contrary, the rail corridor, as a linear asset, acts like a pipeline. A disruption of just a small portion of the rail corridor creates a dramatic impact throughout the asset [3]. In areas of moderate traffic, a single problem in a 2-track configuration can cut throughput by 50% or more. In addition, when the track is operating at capacity, disruptions typically result in lost and unrecoverable service revenue.

3 System for Prioritized Rail Corridor Asset Management

A system for Prioritized Rail Corridor Asset Management is comprised of two integrated information management systems (Figure 1). Optram's ORIM Railway Infrastructure Management (ORIM) system uses condition, traffic, and work data to develop a prioritized work plan. The prioritized plan is sent to MROI's MAXIMO enterprise asset management system to manage and monitor the maintenance-of-way work.

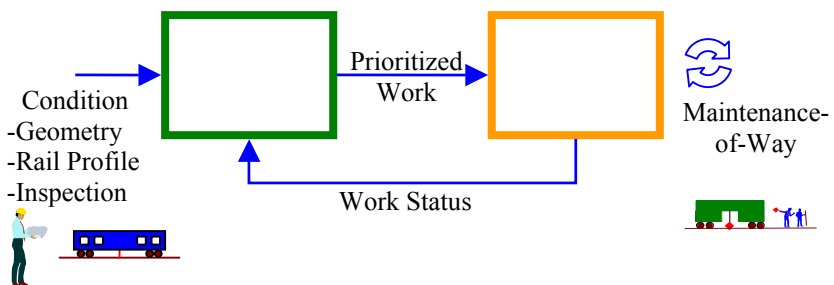


Figure 1: Prioritized Rail Corridor Asset Management.

3.1 Optram's Railway Infrastructure Management (ORIM) System for Prioritization

Optram's Railway Infrastructure Management (ORIM) System is designed to analyze and correlate multiple data-sets about the rail corridor to produce a



prioritized plan of maintenance and capital corridor work. Optram, in collaboration with Amtrak, has been continuously developing this prioritized approach since 1997. Over that time Amtrak has continuously lead in the development and adoption of technology and solutions that incorporate prioritized rail asset management in Amtrak's system called *Infrastructure Information Management System* [4].

ORIM uses existing data about rail corridor configuration, condition, usage and maintenance to develop a prioritized plan. The rail corridor configuration includes the position and relation of assets along the right-of-way. Examples of condition data include track geometry, rail profile, rail defects, track exceptions, and visual inspections. Usage data examples include dynamic and static loads, wheel condition, and train movements. Work includes type of labor, materials and equipment use to maintain the right-of-way.

The Track Maintenance Management White Paper [3] paper describes ORIM as a Track Maintenance Management (TMM) system. It describes TMM systems as assisting rail organizations developing answers to the fundamental question of "What should I work on and why?" TMM systems collect and analyze condition, usage, work and performance data and relate the results to the assets located in the right-of-way. Performance data is derived from the degradation of condition over time and the effectiveness of repair work.

Optram's ORIM integrates existing rail corridor data sources and aligns it with a linear model of the track network. This integrated model is then analyzed to develop a prioritized work plan, to include near term and longer term maintenance work, including capital planning. As the work flow progresses, the status and performance of the work is fed back to ORIM to again refine the prioritized plan.

When freight and passenger rail organizations use Optram's Railway Infrastructure Management (ORIM) system they are able to shift their day-to-day work concentration:

- From generalized, reactive repairs to proactive focused repairs
- From annual production cycles to TMM-enabled targeted programs

Both approaches address root-causes, significantly reducing the expense of repeated repairs.

Heide et al. [1] showed in a substructure application of TMM that without additional capital investment, a 10% annual savings in maintenance expense was possible by better targeting maintenance operations, and deploying resources to go only where and when they are needed [1]. Higher returns were realized when capital investments were made in the infrastructure.

3.1.1 ORIM's return-on-investment, Scotland case study

During the summer of 2002 a 3-month project was undertaken of the Optram Right-of-way Infrastructure Management (ORIM) product with an aim of demonstrating the potential for 10% or more improvement of maintenance efficiencies. In 2004 the pilot project covered 47 route miles of a double-track



commuter passenger rail line. In 2003 the project was extended to include data for a 58-mile section of an intercity main line track.

The areas of track studied under the project were high priority lines that struggle with expensive track related train delays. The delay penalties mount up quickly considering that a 15-minute delay of a single train results in a penalty of £2670. With the possible advent of an additional 30 trains per day on the line, similar delays could amount to daily delay penalties in excess of £80,000.

Working with the customer, the project team identified an initial list of 10 areas where ORIM can be used to reduce maintenance costs and time delays. From this list the team selected three areas to show reduced maintenance costs and time delays:

1. Areas of rough track geometry,
2. Areas of repeat rail defects, and
3. Information-sharing and communication.

These three areas were selected because they represent: a range of savings-types, data was available within the project time and budget constraints and they provided edifying examples of the potential savings obtainable through ORIM use.

The project results concluded by using 3 of the 10 savings areas over the 1000 km of track the system would provide:

1. a strategic advantage for the maintenance contractor through delivery of improved reliability,
2. saving at least £10M per year through improved understanding and control of track infrastructure, more effective maintenance, reduced delay penalties and improved work window productivity [2].

3.2 MROI's Enterprise Asset Management System

MROI's MAXIMO system provides a unified approach to managing all classes of assets, across the enterprise. MAXIMO provides real-time life cycle asset management, from sourcing through final disposal of assets. In addition, MAXIMO includes management of the work flow to maintain the assets.

Once the ORIM "engineering assessment" of asset condition and historical performance is complete, MAXIMO manages the logistics of assigning the steps to complete the work.

4 Modelling the Rail Corridor

ORIM develops its prioritized plan by relating multiple data sources about the rail corridor. The linear characteristic of the corridor requires a unique model to manage and interpret the continuous nature of related information. ORIM correlates data by the relative position within the corridor but without regard to discreet features. This type of data relation called "*continuous linear relation*" and is in contrast to the more typical relation of data by discrete "*segmenting*".

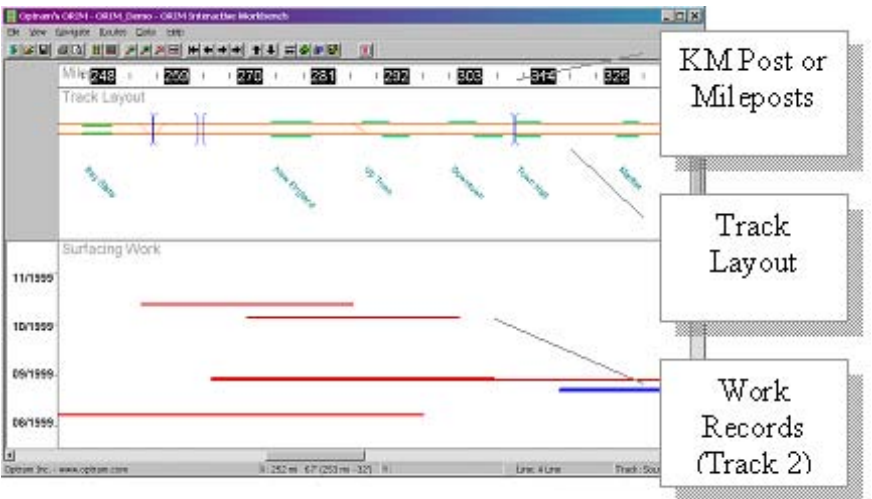
Segmenting of the track network breaks the rail network into discrete segments where data is grouped into buckets with areas defined by corridor



features (e.g., tangent and curve track, bridge transitions, areas of common track components) or buckets defined by even sizes (1/8 mile segments, 1-km segments). Segmenting often groups data that should not be correlated and thus blurs the interpretation of that data.

Figure 2 is an example ORIM track layout and work diagram. The figure shows work performed on the “South” track, which is represented by horizontal lines on the bottom half of the figure. The horizontal axis indicates where the work starts and ends (marked by the km/Mile Post markers on the top of the figure). The vertical axis represents when the work is performed (marked by left hand date scale showing month and year).

It is apparent that work is performed regardless of the placement of such features as switches and bridges.



ORIM is protected under US patent 5,978,717

Figure 2: Rail Corridor Assets Model.

ORIM’s use of “*continuous linear relation*” provides the ability to:

- Accurately pinpoint problem locations
- Avoid blurring of unrelated data
- Interpret multiple issues from a common data set
- Add new linear data sets without re-segmenting

5 Technology to prioritize

To create a prioritized plan the following elements are needed:

- A system that captures and integrates the flood of data collected on a rail corridor.
- A system that schedules, assigns, combines workflow and monitors the people, equipment, and material to complete the asset maintenance and



renewal work. This includes where on the linear asset the work is performed.

- A simple and efficient presentation of portraying the large data set as information.

To provide this solution Optram and MROI combine Optram's linear asset management information system with MROI's strategic asset management information system. These elements combined with effective execution can turn all the data into a prioritized scheme for optimizing resources to maintain the highest throughput, reliability, and safety.

6 ORIM – MAXIMO Integration

Together, ORIM and MAXIMO systems provide:

- Targeted, proactive maintenance
- Accurate long-term maintenance planning based upon actual asset condition, work history and performance
- Automatic generation of pre-planned work orders and associated logistic support elements

Close integration of ORIM and MAXIMO systems also streamlines the asset analysis/work-order maintenance process, improves user productivity, reduces potential errors between these two critical functions, and lowers railroad operators' overall IT support costs.

Asset, condition indicators and work-order data is exchanged continuously between the ORIM and MAXIMO to facilitate a focused and streamlined Track Maintenance Management process. Further discussion of the ORIM and MAXIMO relation is described in [3].

The ORIM / MAXIMO integration streamlines the asset analysis and work-order maintenance lifecycle process, improving user productivity, reducing potential errors between these two critical functions, and lowering railroad operators' overall IT support costs.

7 Conclusion

Managing a reliable and safe rail corridor is challenge. Having simple and readily available information to prioritize the use of limited resources (people, material, equipment, and work windows) can dramatically affect the rail corridor reliability, safety and profitability. Using a Prioritized Rail Corridor Management system can convert the vast, continuous and varied stream of rail corridor data into management intelligence required to continuously manage the safest, least expensive and most reliable rail corridor under constrained resources.



References

- [1] Walter Heide, Conrad Ruppert, Ernest T. Selig and Willem Ebersöhn, Engineering Approach to Track Substructure Management on Conrail and Amtrak, *AREA Annual Meeting presentation*, March 26, 1996.
- [2] “Scotland Alliance ORIM Pilot Project Report, A Summary Report for the Railtrack and First Engineering Scotland Alliance”, May 6, 2003
- [3] “Track Maintenance Management Executive White Paper”, Optram Inc., May 2003, www.optram.com
- [4] Dr. Willem Ebersöhn, Conrad J. Ruppert, Jr., “Railway Infrastructure Maintenance Management at Amtrak”, UK Institute of Mechanical Engineers, September 2002.
- [5] ORIM is protected under patent 5,978,717



Section 2

Safety

This page intentionally left blank

An assessment of hazard probability due to Pentium processor errata in automatic train control applications

C. Bantin

Alcatel, Canada

Abstract

The Alcatel automatic train control products make use of a single-board computer that has been designed specifically for the railway environment. The computer is based on the Pentium processor and is extensively used for automatic train operation and automatic train protection functions. It is also used as a security authentication gateway interfacing to the radio-based data communications system.

The development of an automatic train control system must be accompanied by a detailed and extensive safety case in order to demonstrate that the required safety integrity level can be obtained. For the Pentium-based processor, the safety case must include the occurrences of errata, or faults in the design and implementation of the processor that are not discovered at the time of manufacture. It may be argued that, since errata are design and manufacturing errors, they are systematic. However, because of the way these faults manifest themselves it could be argued they are random. In fact, for any one processor, there is a random errata discovery process based on the fact that all the processors in use are operating simultaneously with different applications and/or different data. There is a particular probability that one of them will discover a fault, or errata. A statistical model is developed based on an in-depth analysis made of Pentium errata, and assumptions about the number of processors in use over the time period of the analysis. A probability is calculated that previously undiscovered errata will be found in one of the processors in an ATC system, and it is demonstrated to become lower than the required hazard probability well before any such ATC system containing these processors is out into revenue service.



1 Introduction

The EN 50129 standard for safety related electronic systems classifies failure integrity (faults) as either systematic or random. The systematic part is considered non-quantifiable and there are procedures for developing a safety case under these conditions. When it comes to microprocessor faults, it may be argued that they are design and manufacturing errors, and therefore they are systematic. However, because of the way these faults manifest themselves it could also be argued they are random. For example, although human error may be at the root of the problem, it is not reasonable to assume they could have been systematically eliminated ahead of time. Therefore, we are left with a random discovery process based on the fact that all the processors in use are operating simultaneously with different applications and different data. There is a particular probability that one of them will discover a fault, or errata.

2 List of major assumptions

- i. Errata present in the Pentium could not reasonably have been systematically eliminated ahead of time. Errata have to be discovered and the discovery process is assumed to be random and therefore amenable to statistical modeling.
- ii. The rate of discovery of errata follows a constant failure rate model, and is independent of any changes that are made as a result of discovering previous errata.
- iii. The number of processors in operation follows a normal distribution.
- iv. The probability of discovering errata is directly proportional to the number of processors in operation.
- v. With reference to [2], all *fault types* must be considered. Only 83.3% of the *configuration types* are considered. Only the 89.1% under the *normal operational mode* are valid. All of the *dependency* events will be considered. Only the 31.9% *lesser severity* events are considered, because the *denial of service*, *hang*, and *crash* events do not result in a hazard threat.
- vi. The categories are orthogonal.
- vii. The rate of discovery of new errata during each month applies uniformly over that month.
- viii. There are 10 million Pentium II processors in operation at the present time, all of which are operating independently, and there is a uniform probability that any one of these processors could discover errata.
- ix. There are 85 TAS platforms in a given ATC system.

3 Main hazard

The main hazard can be identified as **action taken on an unintended telegram** that exposes to the system to a safety-related incident. In this case the hazards are caused by threats from undiscovered errata in the operation of the Pentium



processors in the TAS Platform. The target value for the probability of the main hazard is 10^{-9} per train-hour of operation.

4 Threats from the processor

The main hazard will result from threats imposed by undiscovered errata in the processor. Therefore, the probability of the hazard existing amounts to calculating the probability that the threats exist, as modified by the defense mechanisms available to mitigate this threat. The following analysis derives a model of the process, which demonstrates that the probability of threats from the errata is within the target value.

5 Errata probability models

There are a number of failure rate models that apply to hardware failures in electronics equipment [1]. These range from a simple constant-rate model to more sophisticated models that take into account repairs and upgrades to the equipment. The situation with processor errata, however, is that they are not simply hardware failures that occur as a result of normal operation. Errata have to be uncovered by events that trigger them through the operation of the processor itself in an otherwise perfectly normal manner, using a wide range of data input. Nevertheless, we can base an errata model on the normal failure rate models.

5.1 Constant rate model

The constant failure rate model assumes that the probability of a failure is constant throughout the lifetime of the equipment and that it does not change as a result of any action taken to improve the equipment. The failure rate is given by,

$$h = \frac{1}{\lambda}, \quad (1)$$

where λ is commonly called the mean time to failure (MTTF). The probability density function over time for the constant-rate model is given by,

$$p(t) = \frac{1}{\lambda} e^{-t/\lambda}. \quad (2)$$

The corresponding cumulative distribution function is given by,

$$P(t) = \int_0^t p(t') dt' = 1 - e^{-t/\lambda}. \quad (3)$$



These functions are shown in figures 1 and 2. The probability of a device failing is greatest at the start of its life and diminishes as time goes on. This is a reasonable assumption for the discovery of processor errata, but it assumes a constant sample size, that is to say the failures are not a function of the number of units in operation. On the other hand it is reasonable to assume that the rate of discovery of processor errata is proportional to the number of processors in operation, therefore, the constant-rate model cannot be used directly.

5.2 Model with learning

Another model for equipment failure is known as the Weibull model and it accounts for the fact that the failure rate may change over the lifetime of the device because of improvements made as a result of previous failures (e.g. changing the operating environment to reduce the number of failures). The parameter β is introduced to account for the reduced, or even increased, rate of failures. Thus,

$$h(t) = \frac{\beta}{\lambda} \left(\frac{t}{\lambda}\right)^{\beta-1}, \quad p(t) = \frac{\beta}{\lambda} \left(\frac{t}{\lambda}\right)^{\beta-1} e^{-(t/\lambda)^\beta} \tag{4}$$

and

$$P(t) = 1 - e^{-(t/\lambda)^\beta}. \tag{5}$$

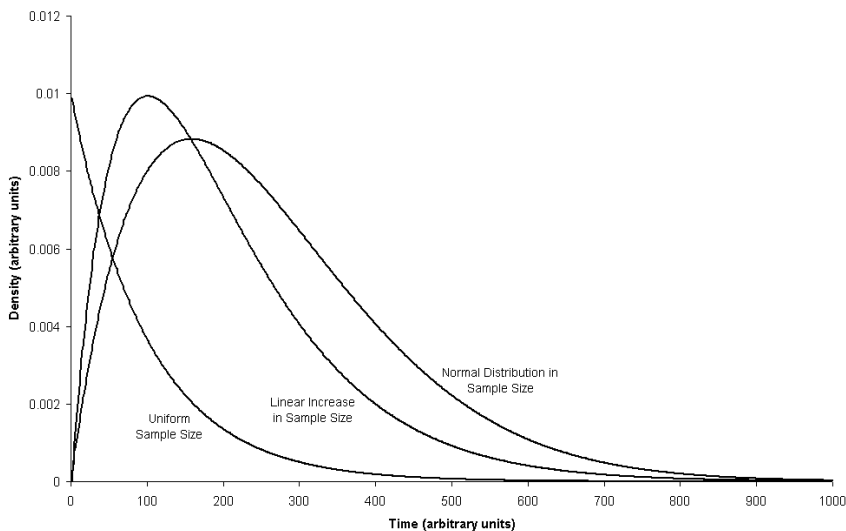


Figure 1: Probability density distribution of errata discovery.



The Weibull model is not really an appropriate model for processor errata because it is reasonable to assume that the rate of failure, meaning the rate of discovery of faults, is independent of any changes that are made as a result of discovering previous failures. The value of β is typically between 0.95 (improvement) and 1.05 (worsening). A value of 1.0 gives the constant rate model.

5.3 Model with linear sample size

In order to have a realistic errata discovery rate model we need to combine the constant rate model with a reasonable model of the number of processors in operation. If we assume a linear growth in the number of units contributing to the process of discovering errata, then we can write,

$$n = Nt, \quad (6)$$

where n is the number of processors and N is the rate of growth. There is a tacit assumption here that there is a limit reached at some time, beyond which there are no further units (or at least some other model applies). We now assume that the probability of discovering errata is directly proportional to the number of units contributing, therefore,

$$p(t) = \frac{N}{\lambda} t e^{-t/\lambda}, \quad (7)$$

$$P(t) = N \int_0^t (t/\lambda) e^{-t'/\lambda} dt' = N \lambda \int_0^{t/\lambda} t' e^{-t'} dt' = N \left(1 - \left(1 + \frac{t}{\lambda} \right) e^{-t/\lambda} \right). \quad (8)$$

The results are also shown in figures 1 and 2. The rate of discovery early in the device life is clearly reduced because of the small number of units in operation. However, the model is still not very realistic because there is no limit to the number of units in operation and it is not reasonable to assume the rate of sales of Pentium processors was in fact linear.

5.4 Model with realistic sample size

A more realistic model for the number of processors in use is to assume a normal distribution where there is a small number at first, a large number during the mature period of sales, and a small number again at the end of the production life. Such a model is given by,

$$n = N e^{-((t-t_o)/\tau)^2}, \quad (9)$$

where $\tau = \sqrt{2}\sigma$ and σ is the standard deviation, t_o is the mean and N is a normalizing factor (depending on the total units sold). Then,



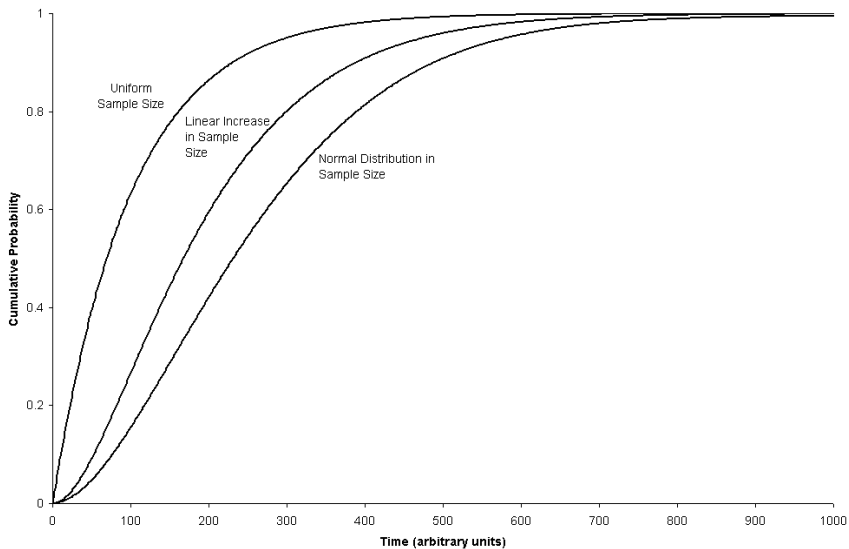


Figure 2: Cumulative distribution of errata discovery.

$$p(t) = \frac{N}{\lambda} e^{-((t-t_0)/\tau)^2} e^{-t/\lambda}, \tag{10}$$

$$P(t) = \frac{N}{\lambda} \int_0^t e^{-((t'-t_0)/\tau)^2} e^{-t'/\lambda} dt'. \tag{11}$$

Figure 3 illustrates the cumulative distribution of processors in use. In addition, figures 1 and 2 show how the new model is more reasonable and we will show next that it is a very good fit to actual errata rate data.

6 Processor errata data

The behaviour of the Pentium II processor up to April 1999 is examined in detail in a study of microprocessor entomology [2], and its application to high-confidence computing systems [3]. Not all of the errata discovered for the Pentium II processor are threats to the main hazard, and the results reported in [2] are used here as both the source of data for the rate of discovery of errata and to determine whether the errata discovered are of the type that can cause a hazard.

6.1 Errata discovered to date

The cumulative distribution of errata reported over the period May 1997 to April 1999, taken from [2], is shown in figure 4. There are a total of 73 errata, and it should be noted that 27 of them were reported immediately.

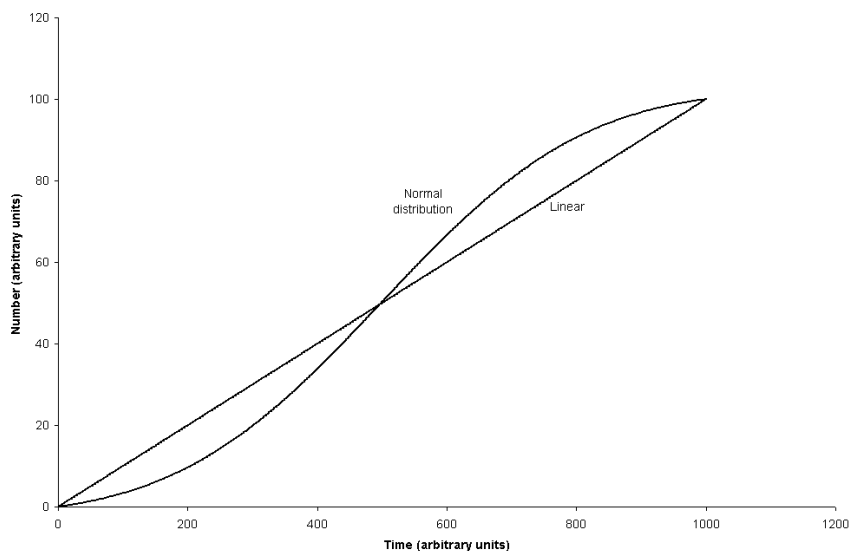


Figure 3: Cumulative distribution of units in operation.

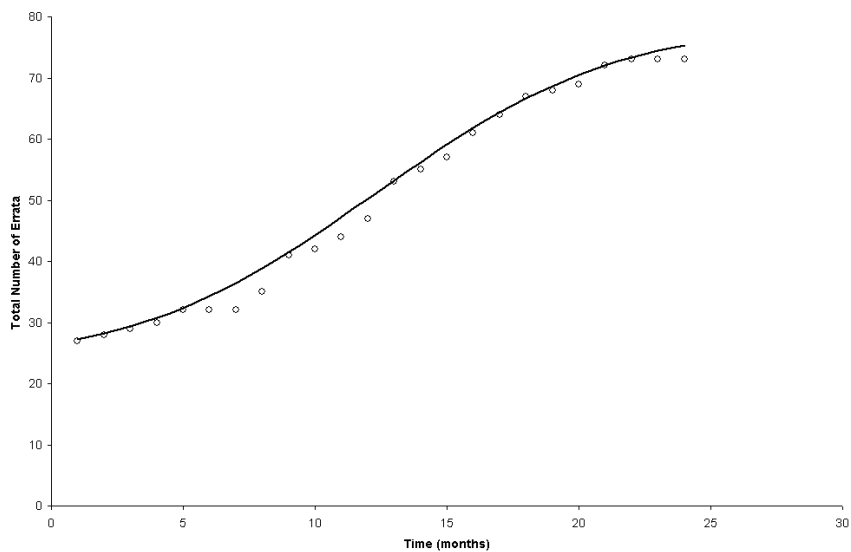


Figure 4: Errata reported from May, 1997 to April, 1999.

6.2 Model fit

A curve of the type described in 5.4 is shown in figure 4 fitted to the data. Note that a worst-case assumption is made in fitting the curve to the upper data



samples on the graph. Also the initial 27 are treated as a delta function in the density distribution and show up as an offset in the curve. The parameters that provide a good fit are as follows:

$N = 3.3$, which is the maximum number of errata reported in a month,
 $t_0 = 13$, which is the mean in months of units in operation,
 $\tau = 10$, which corresponds to a standard deviation of 7.1 months for units in operation,
 $\lambda = 15$, which is the “mean time to fault discovery” in months,
 Offset = 26.5, which accounts for the initial delta function.

This model predicts that there should have been 75.26 reported errata by April 1999, as opposed to the actual value of 73.

6.3 Expected future errata

Using the curve fitted to the data we can calculate the maximum expected number of errata for the Pentium II throughout its lifetime. To do this we simply extend the curve to very large values of time. The result is 77.86. This means that 2.6 more errata are to be discovered during the remaining lifetime of the Pentium II processor (since April 1999). As a worst case, however, we assume that there are $77.86 - 73 = 5$ (rounded up to the nearest integer) more errata to be found than have currently been reported. Note, that by 2006 virtually all 5 should have been found, but since there are no records beyond April 1999 we must assume that all 5 are still waiting to be found. We will discuss this assumption further later on.

6.4 Classification of errata

The data in [2] is also presented by the classification of errata. The classifications used are *fault type*, *configuration*, *operational mode*, *dependency* and *severity*.

We will assume here that all *fault types* must be considered. For the *configuration*, only the 83.3% of the errata under the *uniprocessor configuration* are valid candidates. The *multiprocessor* and *F.R. checking* categories are not valid since the TAS Pentium IIs are not operated in these modes while doing ATC functions. Similarly, only the 89.1% found under the *normal operational mode* are valid, as opposed to *starting*, *test*, *recovery*, and *system management* modes. All of the *dependency* events will be considered, but only the 31.9% *lesser severity* events are considered. This is because the *denial of service*, *hang*, and *crash* events do not result in a hazard threat.

If we assume these are independent (orthogonal) categories then the fraction of the new errata to be found that could be a threat can be computed as,

$$F = 1.0 \times 0.833 \times 0.819 \times 1.0 \times 0.319 = 0.218 \text{ or } 21.8\%. \quad (12)$$

The data is further categorized according to where the errata occurred in the processor hardware. Of the errata, 50.5% were in the performance *delivery architecture*, and are valid candidates, while 49.6% were in the *confidence*



assurance architecture and therefore do not present a threat. Therefore, we can modify the fraction by the factor 0.505,

$$F = 0.505 \times 0.218 = 0.119 \text{ or } 12\%. \quad (13)$$

In summary, based on historical data, only 12% of the errata could have potentially caused a threat.

7 Threat model based on non-redundant processors

This section calculates the probability that a threat exists assuming that there is no advantage to be gained from the 2-out-of-3 (2oo3) processor redundancy that is implemented on the TAS Platform. The probability is based entirely on the above model derived to fit to the reported data [1].

7.1 Target hazard rate

The target probability for the main hazard is,

$$P_h = 10^{-9} \text{ per train hour.} \quad (14)$$

In the context of this analysis we interpret this to mean that there is a probability of P_h that a particular train will experience a hazard in any one hour period due to a threat from undiscovered errata in the Pentium II processors associated with that train during the hour.

7.2 Existing errata with no fix

There are 44 known errata that are listed as having “no fix” [2]. In the context of this analysis this means that there are 44 known errata that could be triggered by the Pentium processors. We consider known errata to be faults of a systematic nature. These are not amenable to statistical analysis and the defenses against the threats posed by such faults are evaluated qualitatively. None of the errata reported to date for the Pentium II, which do not have a fix, impact the TAS platform in a safety critical manner.

7.3 New errata

According to the model developed in section 5, we assume there are 5 errata yet to be discovered in the lifetime of the Pentium II.

7.4 Errata causing a threat

Of the errata yet to be discovered, only the fraction $F = .12$ will cause a threat to the main hazard, according to the classifications in [1]. Therefore the number of threatening errata yet to be discovered is,

$$N_e = 5 \times F = 5 \times 0.12 = 0.6. \quad (15)$$



7.5 Probability of discovering a threat per operating hour

In the interim period from April 1999 to July 2006 no errata have been identified. It does not matter whether this is because they were not discovered, not reported or simply that the records were not updated. Our model assumes they are yet to be discovered, and the potential threat of concern is they will be discovered by one of the TAS platform Pentiums in an ATC project, given all the devices in operation at the time. What this means for our model is that we can remove a block of time from April 1999 to, say, July 2006 and continue the probability model at the later time. We can then calculate the probability of errata being discovered during any one hour over the time period of interest starting in July 2006 (or another suitable starting date).

According to our model of the errata discovery process, the rate of discovery is a maximum at the start of the time period and decreases with each succeeding month. The value for the first month is 0.678 errata per month. If we assume this applies uniformly over the month then the rate per hour is $0.678 / (30 \times 24) = 0.000942$. The rate after 5 years is virtually zero (less than 10^{-27}). This means that the worst-case rate of discovery of errata that will cause a threat is,

$$P_e = 0.6 \times 0.000942 = 0.000565 \quad (16)$$

per hour, and this corresponds to the first hour of operation beyond April 1999.

7.6 Probability of threat per train processor

Our model includes all of the Pentium II processors operating in the world, including those in a particular ATC project. These processors are operating on either different applications or different data on similar applications. Either way it is reasonable to assume that all of these processors are operating independently and there is a uniform probability that any one of these processors could discover errata.

Of interest here is the probability that one of the TAS platform Pentiums will discover the errata. We assume there are 85 TAS platform modules planned for a typical ATC system. There are conservatively 10 million Pentium II devices in operation, all of which are performing calculations more or less continually as are the TAS platform Pentiums. Therefore the probability that a TAS platform Pentium II will be the unlucky candidate is,

$$P_{Tp} = 85 / 10^7 = 8.50 \times 10^{-6} . \quad (17)$$

According to the assumptions of this section the operation of the redundant processors in the TAS platform module provides no advantage. There are three Pentium devices configured to detect independent random failures using a 2oo3 majority-voting scheme. At one extreme we can argue that since the three devices operate with identical software on identical data there is a high probability that they would all discover the same errata at the same time and the



redundancy is of no use. Therefore, for this section of the analysis the three devices will be considered as one and the probability of the redundancy failing is by default,

$$P_f = 1.0. \quad (18)$$

We can now calculate the probability of one of the ATC TAS platform devices encountering threatening errata in the first hour of operation as,

$$P_t = P_f P_e P_{TP} = 4.81 \times 10^{-9}. \quad (19)$$

7.7 Probability of hazard

Encountering threatening errata does not necessarily present a hazard. Even if an erroneous telegram is transmitted, and there are no errors in transmission, the received telegram must undergo some checks before it is accepted. Of the many checks performed we cannot rely on any of those involving data appended to the telegram after it is generated. These include sequence numbers, CRC checksums, authentication hashes or encryption. The only candidates for protection against a threatening telegram are the persistency and consistency checks. A threat can only become a hazard if these checks fail.

Again, according to the assumptions of this section, the operation of these telegram checks, at this time, is assumed to provide no advantage. Therefore, the probability of the checks failing on all three Pentium devices is by default,

$$P_{cf} = 1.0. \quad (20)$$

The final calculation for the probability of a hazard during the first hour of operation is then,

$$P_h = P_t P_{cf} = 4.81 \times 10^{-9}. \quad (21)$$

Using the model for errata discovery we can plot the probability over time as shown in figure 5. Note that the target of 10^{-9} is met after 5 months of operation. This will cover the period of testing before the system goes into operational service.

8 Conclusions

We have defined a main hazard as action taken on an unintended telegram, and mathematically modeled the probability of occurrence due to undiscovered Pentium processor errata that are a threat to the system. We have used standard probability distribution functions with parameters selected by fitting the model to actual field data in order to predict the number of threatening errata yet to be



discovered. With worst case assumptions the results indicate that the probability of a hazard meets the target of 10^{-9} per train hour of operation within 5 months from power-up (start of computing), a period of time which will easily be covered by system testing and commissioning and long before the system goes into revenue service.

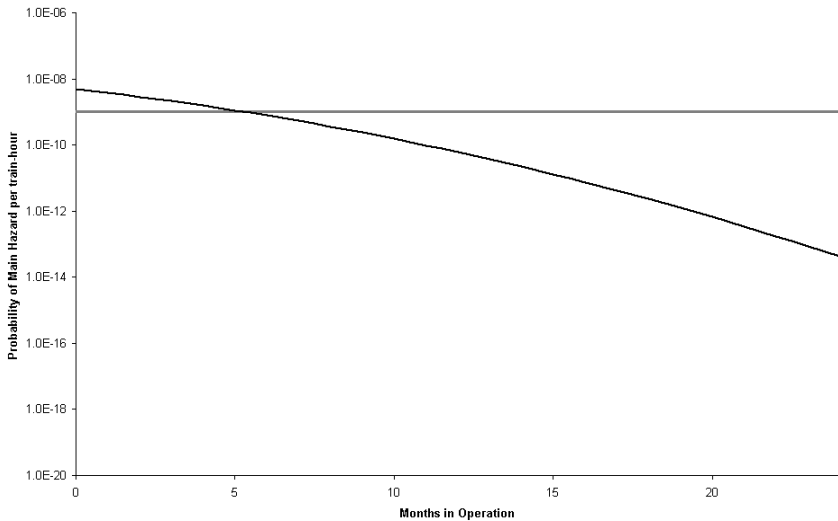


Figure 5: Probability of occurrence of the main hazard per train-hour of operation based on only the rate of discovery of errata in the Pentium II processor.

9 List of variables

h = failure rate ($1/h$ = MTBF)

$p(t)$ = probability density function

$P(t)$ = cumulative probability function

β = learning factor

n = number of processors in operation

N = rate of increase in the number of processors in operation

t_0 = mean (in months) of units in operation

$\tau = \sqrt{2\sigma}$

σ = standard deviation (in months) for units in operation

λ = mean time to fault discovery (in months)

F = fraction of errata that are threatening

P_h = main hazard probability (per train hour)

N_e = number of threatening errata to be found

P_e = probability of a processor discovering errata

P_{TP} = probability of a TAS platform processor being involved



P_f = probability of redundancy process failing

P_t = probability of TAS platform processor discovering threatening errata

P_{cf} = probability of data checking process failing

References

- [1] Zang, H., Cutright, E., Giras, T., Time-varying failure rate for system reliability analysis in large scale railway risk assessment simulation, Computers in Railways IX, WIT press, 2004, pp313-322.
- [2] Avizienis, A., He, Y., Microprocessor Entomology: A Taxonomy of Design Faults in COTS Microprocessors, Dependable Computing for Critical Applications 7, 1999, 6-8 Jan 1999, pp 120-121.
- [3] He, Y., Avizienis, A., Assessment of the Applicability of COTS Microprocessors in High-Confidence Computing Systems: A Case Study, Proc. International Conference on Dependable Systems and Networks, 25-28 June 2000, pp 492-500.



This page intentionally left blank

Role of supervision systems in railway safety

F. Belmonte, J.-L. Boulanger, W. Schön & K. Berkani
*Laboratoire Heudiasyc, Université de Technologie de Compiègne,
 France*

Abstract

The new generation of supervision systems in industry can achieve operation from display process variables to all automated control where the human is just the monitoring automaton. In the railway specific industry, supervision is organised in switching zones and aims to be centralised in an Integrated Control Centre. Such centres implement integrated and computer based systems that perform train protection, train operation and supervision. Thus railway dispatchers using supervision have their tasks considerably simplified. Although considered today as not safety critical, railway supervision systems can contribute to safety in some scenarios where an appropriate decision of a supervision operator could notably reduce the severity of accidents. That is in particular the case for residual scenarios (intervention of maintenance teams on the tracks, manual operation of trains not protected by train protection system, coupling/uncoupling, emergency requiring the stopping and evacuation of a train etc) only covered by procedure, thus requiring human intervention by a person supposed correctly informed on the state of the system, thanks to the data provided by the supervision system.

Keywords: supervision, automatic train control, safety, human factor.

1 Introduction

Whatever the degree of human presence aboard trains, all railway lines have however a system which centralizes for operators of a central room, as well the operations of traffic control (signalization, traced routes etc.) as operations of traffic regulation. Formerly analogical, today technology is becoming numerical. Therefore, it is now possible to generalize the centralization of many functionalities formerly carried out locally by operators on-site (signals and



switching devices handling, traction energy management, disturbance management, public places monitoring). This technological development improvement must be accompanied by an evolution of operator's culture. On the one hand, the increasingly frequent use of numerical technology in safety functions requires rigorous process development. On the other hand, the activity of human operators changed much: operators on-site (drivers, pointsmen) are becoming more and more supervisors of automats. They must apprehend, analyze the remote situation and as far as possible often act with the help of remote controls.

This paper is structured as follows. Section 2 presents a general overview of supervisory control, the definition of supervision of railway network, and how the supervision systems are implemented and deployed. Section 3 deals with safety concepts and implementation in automatic train control systems. We explain how safety could be improved with the help of supervision systems. Section 4 presents our supervision platform which is installed in University of Compiègne. Human behaviour experiment will be performed using this platform.

2 Supervisory control/command theory

Industrial process supervision covers a wide range of applications, from simple control to complex and heterogeneous system control. The aim of this section is to introduce basic definitions about supervisory control system. They define the context of our approach.

2.1 Definitions

Originally, the term supervision refers to the human interaction between a superior and his subordinates. Dictionary [8] defines supervision as monitoring and overseeing a task without going into particulars.

In control and automatic context the definition of supervisory control is the control and the monitoring of a system including, where appropriate, those operations which ensure reliability and safety. In addition, "Automation dictionary" gives the following meanings to supervision control:

1. "Control in which the control loops operates independently subject to intermittent corrective action". For example, the corrective action can be set point changes by an operator or another external source;
2. "A term used to imply that a controller output or program output is used as an input to other program, e.g., generation of set points in cascade control systems", used to distinguish from direct digital control.

There are two major models of supervision from scientific literature. Supervisory control theory is the first of them. This model was introduced by Ramadge and Wonham [4] from automatic theory domain. It consists of analyzing discrete events control systems in a formal manner. Railway Traffic could be reduce to discrete event control system. The process is modelled as a state-transition structure, labelled a Finite State Machine, in which transitions are labelled as controllable and uncontrollable. Control actions are exerted upon the



process in feedback mode by another FSM termed supervisor that disables controllable transitions in order to satisfy given behaviour specifications. In other words the supervisor disables process controllable transitions in such a way that it restricts the process behaviour to a subset of all the possible trajectories.

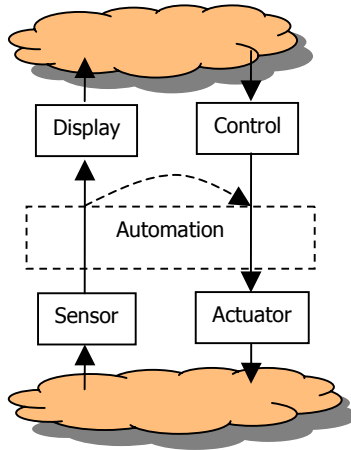


Figure 1: Sheridan supervisory control model.

Another model, even more generic was introduced by Sheridan [1], his concept of human supervisory control is depicted in Figure 1: by five main tasks. In supervisory control, a human operator monitors a complex system and intermittently executes some level of control on a process, always acting through some automated agent. During supervisory control, an operator plans an activity and instructs it to a computer. A human checks if the actions are executed and intervenes in case of computer failures. Finally, the human learns from the experience. The interrelationships between the five supervisory tasks are characterized by three loops (see Figure 2:). The inner loop is the monitoring where observations lead to further detections and diagnostics. The intermediate loop relates operator interventions to reprogramming or modifying the computer system to follow new goals or reach new states. The outer loop feeds the latest process knowledge and experience from learning back to planning. The three loops operate on different time scales. The inner monitoring loop is executed frequently while modifications of process goals (the intermediate loop) are done less frequently. Finally, the overall process plans and goals are changed over relatively long time intervals.

Supervisory control is that a human operator is intermittently giving a command and receiving information from a computer that itself is an automatic controller to a given system [1].

Implementation of supervisory control system are based on SCADA systems [3], the next section presents briefly their principal characteristics.



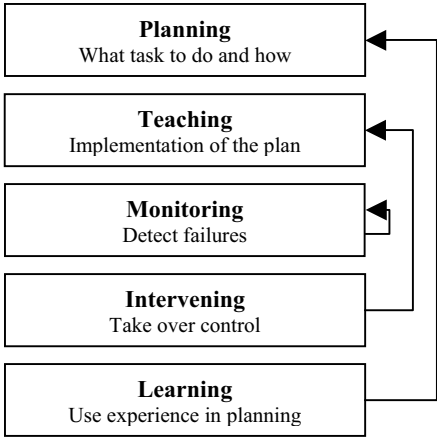


Figure 2: Five tasks in supervision (from Sheridan).

2.2 SCADA

SCADA is an acronym that is formed from the first letters of the term “supervisory control and data acquisition”. SCADA is the implementation of supervisory system.

At the centre of SCADA system, the operator accesses the system by means of an operator interface device. The operator consoles functions as the operator’s window into the process. It consists of a video display unit (VDU) that displays real-time data about the process and a keyboard for imputing the operator’s commands or messages back to the process. Other cursor-positioning devices, such as a trackball, mouse, or touch screen may be used. If the system is very simple, it may be sufficient to have a set of annunciator windows that mimic the condition of the remote process. Often, an audible signal for alarms will be included. The operator interfaces with the master terminal unit (MTU), which is the system controller. It can monitor and control the field even when the operator is not present. It does this by means of a built-in scheduler that can be programmed to repeat instructions at set intervals. MTU must communicate with remote terminal units RTUs that are located away from the central location. A SCADA system may have as few as one RTU or as many as several hundred. RTU communicates with RTUs using the communication network (Radio, serial communication, Ethernet, etc.).

SCADA systems (see Figure 3:) consist of three functional components: communications equipment, remote terminal units and master terminal unit. RTU gathers information from field and centralise this through communication network to MTU. Then MTU treats this information and displays them to operator through Human Machine Interface. MTU also achieves tasks (automation tasks) but always present information (treated or direct from field) to operator.



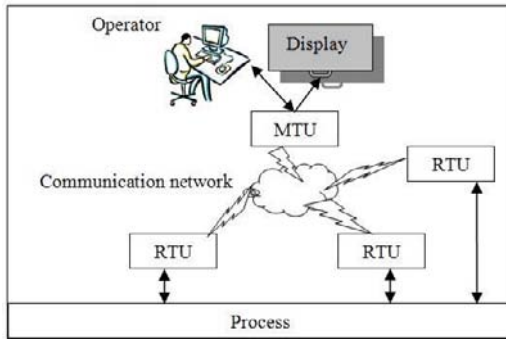


Figure 3: SCADA system.

3 Railway supervision

3.1 Integrated Control Centre

Railway supervision system includes 3 hierarchical levels: (1) Route control level includes field control and commands as route setting, train control and protection functions. (2) Traffic flow control level insures conflict solving, and time graph monitoring functions. Finally (3) the highest level performs management and planning tasks. There are several implementations of the three levels. One could separate each level in three control rooms. This is the case for large railway networks with a lot of inter connexions. These kinds of networks are segmented in switching zones where signal boxes achieve route control level. Generally, centralized traffic control (CTC) centre gathers several switching zones. CTC achieve the level 2 of supervision (traffic flow control). Finally, coordination control centre manages the entire network. It could be international, national or regional organisation. Other implementations centralise the three level of supervision in a unique centralised room called operation control centre (OCC). This is the case for metro and urban railway network. Indeed, such networks are characterized by their unique line and their rare connections with other networks.

These kinds of implementations could be explained by the technology used until today that not permitted to integrate very large network control in a single room. With the emergence of high intensive computerised systems in railway supervision, operational company use more and more integrated control centre (ICC) (see Figure 4:). ICC could gather operation and control on the three levels of supervision. ICC is the generalisation of urban applications OCC of all types of railway supervision. Newest networks as high-speed line in France use an ICC called train-regulating signal box that deliver both level one and two of high-speed line. The third level cannot be integrated in such circumstance because of the existence of few connections with other national switching zone (as specific signal boxes for large stations managements for example). Indeed,

management and planning must be decided in a highest regional or national commandment.

At the same time, Automatic Train Control system (ATC) performs three underlying functions in railway application: protection, driving assistance and supervision. In this system, supervision function is performed at the ICC and the train driver's tasks could be simplified considerably and safety stems from protection.

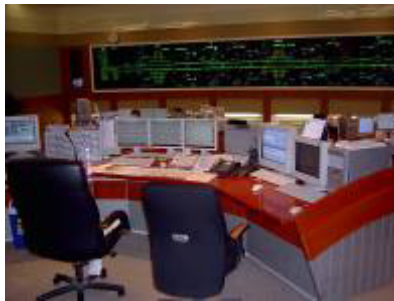


Figure 4: Photography of ICC.

3.2 Automatic Train Control

ATC system is composed of three subsystems giving in decreasing order of underlying safety: Automatic Train Protection, Automatic Train Operation and Automatic Train Supervision. The followings sections present each subsystem.

3.2.1 Automatic Train Protection

Automatic Train Protection system performs a range of functions that protect passengers and equipments against principal dangers related to railway circulation (collisions between trains, collisions with objects, excessive speed, derailments). The line is cut out in sectors called blocks or track sections. Spacing between trains is ensured by the function of block occupancy management. Train detection by track circuits (one track circuit per block) indicates train position and insures trains spacing. The guiding principle of railway safety consists, via side signals indications, to authorize the presence of only one train by section. In the majority of cases, sections are fixed (fixed block) and side signalling consists of traffic lights in limits of each block. For the most recent systems, blocks are moving with the train and their length are evolving with train's speed and way occupation (moving block), which allows an optimization of line's transport capacity. In both cases system ATP is able to detect incompatible speeds in accordance to stop limit distances at the end of sections and at extreme the penetration on occupied sections (and naturally acts as consequence by an emergency braking of the train if one of these events occurs). The usual terminology distinguishes from the ATP the interlocking whose essential function is to prevent incompatible movements of trains on the switching zones.



3.2.2 Automatic Train Operation

Automatic Train Operation or “autopilot” controls fully or partially the trains according to a fixed timetable. Timetables are defined each day to prepare and organize exploitation. ATO manages programmed stops in stations, doors control and the dwell time respect. Then it starts again the train after each stop in case of integral automated system. In partial automation cases, doors opening, doors closure and depart order are given by aboard operator (driver or “attendant”).

3.2.3 Automatic Train Supervision

Automatic Train Supervision insures monitoring of all subsystems which compose the line (trains included) and regulation functions. Each components state are represented on visualisation panels called Schematic Control Panel (SCP) presenting a global view of the line and on the working stations of operators which can provide more or less detailed sights and tables or graphs of telemetries. Information presented can thus be more or less fine according to operational constraints' of the line. The ATS thus makes it possible to apprehend in a total way the state of line operation (by the synoptic in general presented at SCP which presents in real time information as sections occupied, switching points position, energy state, etc.). It can also focus operator attention on particular equipment (fixed or on board a train) via telemetries also updated in real time, allowing alarms visualisation on which operators can obtain details. The ATS also presents data making it possible to ICC operators to carry out traffic flow control (for intercity networks, this function is carried out by a particular operator called regulator) which consists to decide the appropriate actions in case of incident to restore a (possibly degraded) mode of operation. Depending of the detail of the situation, other trains on the line could be delayed in order to manage intervals.

3.3 Safety

CENELEC is a safety reference frame. It is particularly applicable to railway domain and is composed of the following standards: EN 50126 [5], EN 50128 [6] and EN 50129 [7]. With this reference frame, the safety measure can take a value among five possible. The smallest SIL (Safety Integrated Level) is 0 when the evaluated application is not critical. The highest SIL is 4 when the evaluated application is very critical. For example, an application is SIL 4 when many human lives are concerned.

As we explained, safety is ensured by the ATP (Automatic Train Protection) functions and interlocking. At the beginning, all these features were implemented with safety relays and analogical circuits. These implementations use *intrinsic fail safe design* (no single failure or likely combination of failures can lead the system to a less safe state than that before the failure). Since several years, numerical systems are emerged in ICC implementation. These systems are SIL 3-4. Thus, the development of such systems needs rigorous methods like those presented in the safety reference CENELEC. These methods concern as much hardware devices as software ones. Finally, numerical systems have several advantages. They are flexible and they have a smaller size than relays



and analogical circuits. For these reasons, ATP and interlocking implementation with numerical solutions will be generalized in the future.

ATO (Automatic Train Operation) features are built on a complex system control. The main consequence is that SIL 3-4 insurance is not reached for this system. Nevertheless, when ATO failures occur, the ATP is charged to catch up the problem. For example, ATO can emit an order that cause an inopportune acceleration. In this case, as ATP controls safely the speed, it can cause the stopping of the train by emergency braking. Thus, ATC safety approach is built on ATO, which safety designed.

Similarly, ATS is not considered classically as safety critical like ATO. In addition, the ATS conception is based on COTS (Commercial Off the Shelf) components. The main objective is to reduce the manufacturing costs. But, the SIL level of COTS components is practically not established. Moreover, the safety analysis of COTS components is practically impossible because the required data are not provided. Nevertheless, several accidents and incidents analysis clearly showed that although the origin of a safety problem could never be attributed to the ATS, an adapted management of a crisis situation and an adequate decision of a quite well informed supervision operator could seriously reduce accidental scenarios consequences. More precisely ATS features assist the operator to execute procedures adapted to the different scenarios an exploitation mode. And, an important report has been done: more the exploitation mode is degraded, more the ATS role is important to ensure safety.

The present context, underlaid by a legitimate concern for the profitability of the infrastructures, considering the competition of road transport (either transport of passengers or transport of freight), generally leads to operating the lines up to their maximum capacity (at least during certain time slots), which thus provokes a strained flow management, requiring, when an operator's intervention is needed, very short delays. Additionally, the human operator getting used to the automation of some tasks which were his in a recent past, is very little trained to act when those helps are not available to him, cases that are precisely situations of crisis. The result of this is at best a clearly enlarged period before the appropriate reaction, effect which adds itself to the growth of traffic mentioned above. In the worst cases, given that the instructions do not usually consider thoroughly the scenarios in which the information about the state of the system is unavailable or very damaged, some completely inappropriate actions from the human operator can be observed (so as to make the situation even worse). The analysis of such scenarios often shows that the operator, for lack of complete and reliable information about the state of the system, creates for himself a coherent but erroneous mental outline in great hurry, on which his decisions will be based; the tension linked to the crisis making him carrying on regardless of other possible plans compatible with the information available to him.

For all the reasons mentioned above, the possible impact of supervision system on safety must be studied, as much for applications of urban transport as of main lines passengers and freight transport (the last two moreover usually sharing the same infrastructures).



4 SPICA rail platform

Studies on human operator behaviours experiments will be performed at the University of Technology of Compiègne (UTC) using a supervision platform “SPICA-Rail” similar to a real one (Alstom Transport’s supervision product: ICONISTTM), installed in our research center (see photography Figure 5:). This equipment will of course include an “environment simulator” making it possible to do “as if” the experimental platform would be really connected to a railway network. The main interest will be the possibility to re-create in laboratory real accidental scenarios, and to be able to confront human operators with these situations, in order to analyze their comportment and their decisions.



Figure 5: Photography of SPICA rail platform.

This platform is already installed but still in validation process. Such experiments suppose a clear definition of requirements to define the most adapted railway network to analyze contribution of supervision’s process in safety. ICONISTTM product is a generic product of railway supervision based on SCADA. For each project, an instance is defined from generic modules of ICONISTTM and datas defining project properties as for example: track plan, stations, switching points or signalization rules. The ATS system installed in UTC includes three levels of control on traffic: manual automated and optimized. The network simulated is divided into interlocking zones (or switching zone). Each zone disposes of local ATS system where operator could take control supervision locally. The global supervision is centralized in a central ATS which gather all local ATS information and command.

SPICA-Rail could simulate the control and the supervision of an entire network by integrating traffic control functions. It includes a wide range of traffic control and management functions including for example: signalling supervision, route setting, train tracking, train describer and timetable management.

In order to simulate crises situation, personal subject of experimentation behaviour will start from an automatic control level in which all operations have an



objective to comply with an operational schedule (timetable) and network regulation. Experiments consist with increasingly insert disturbance on the network and evaluate human behaviours.

5 Conclusion

Safety development of supervisory software is far from obvious, but we argue that well-understood interfaces between man and machine could contribute in safety. Major problem come from bad requirements. Our purpose intends to analyse human supervisor behaviour in such particular serious cases in order to highlight lack in requirements. Capability for safety and automate working adequacy of human performance are the underlying research plan.

As a conclusion we expect to propose a methodology to develop new supervision systems taking into account human factor from the beginning of the design process. That will make it possible to use a process by *feed forward* (anticipation based on preliminary studies [2]) instead of a process by *feedback*.

References

- [1] Sheridan, T. B., *Telerobotics, Automation, And Human Supervisory Control*, Cambridge, MA:MIT Press, 1992.
- [2] Norman, D.A., *The invisible computer: Why good products can fail, the personal computer is so complex, and information appliances are the solution*, Cambridge MA: The MIT Press., 1998.
- [3] Boyer, S.A. SCADA: *Supervisory Control and Data Acquisition*, ISA-The Instrumentation, Systems, and Automation Society, 2004.
- [4] Ramadge P. J., Wonham W. M., Supervisory control of a class of discrete event process, *SIAM Journal on Control and Optimization*, **25(1)**, pp. 206-230, 1987.
- [5] CENELEC EN 50126: Applications ferroviaires – Spécification et démonstration de Fiabilité, Disponibilité, Maintenabilité et Sécurité (FMDS), 1999.
- [6] CENELEC EN 50128: Applications ferroviaires – Système de signalisation, de télécommunication et de traitement – Logiciels pour systèmes de commande et de protection ferroviaire, (2001).
- [7] CENELEC EN 50129: Applications ferroviaires – Système de signalisation, de télécommunication et de traitements – Systèmes électroniques relatifs à la sécurité pour la signalisation, (2001).
- [8] *Webster's Encyclopaedic Unabridged Dictionary of the English Language*, Grammery Books.



Automatic train controller safety simulation

R. A. V. Gimenes, J. R. de Almeida Jr. & T. R. Nogueira

Safety Analysis Group – Escola Politécnica da Universidade de São Paulo, Brazil

Abstract

Nowadays, we are observing an increased demand for better and safer mass transport systems. The supervision and control of these systems is made through an architecture known as Automatic Train Controller (ATC). The use of processors in ATC provides new challenges in a safety analysis. A typical railway system has a Speed-Distance Profile Generator that determines the maximum allowed speed in each track circuit. Dangerous situations are verified through the relative positioning and speeds between trains, switching machine positioning and other restrictions from operational commands generated by the ATC. Independent Safety Auditors should consider the failure modes of hardware and software in use. The safety analysis should consider alternative techniques to complete the complex task of evaluating how safe is the use of the processor and its associated software. Therefore, the use of simulation can improve and increase safety analysis, searching for fault states that could not be found in a static analysis. The main goal of this paper is to describe the development of a tool that simulates the behaviour of trains' movement in a subway system, with boolean expressions. The set of boolean expressions coordinates all the movements in a subway line and the simulation provides the possibility to find out lack of safety, considering different combinations in those boolean expressions. Another important goal is to simulate equipment faults in order to investigate problems not visible in a static analysis or even in a practical field test. Preliminary results have shown that the use of a simulator to execute boolean expressions offers a great variety of tests, allowing the detection of unsafe situations, complementing software tests validation in a final release. Through simulation, it is possible to observe the behaviour of simulated objects in specific internal points which improves the completeness in safety analysis.

Keywords: railway system, environment, safety, simulation, automation, modelling, boolean expression, microprocessor.



1 Introduction

Nowadays, we are observing an increased demand for better, safer and more available mass transport systems, inclusive in the railway systems. The supervision and control of these systems is made through an architecture known as Automatic Train Controller (ATC).

The main ATC System goal is to automate the train circulation and to obey the operational commands with all safety requirements. The rails that compose the track are divided in track circuits. So, one of the main functions of the ATC is to send speed codes to the track circuits and, at the same time, to verify if each track circuit is occupied by a train.

Other components of a railway system are the gates and the switching machines. The Gates are important to control the entrance in an Interlocking Location, which contain switching machines that coordinate the train crossing through different track circuits.

The Track Circuit equipment informs the ATC if there is a train within it. The ATC also receives information from the switching machines position and others ATCs. Those information shows to ATC snapshots from the railway, providing capacity to determine changes about the maximum speed allowed for all Track Circuits, switching machines position and gates states.

The railway has a Speed-Distance Profile that determines the maximum allowed speed in each track circuit. The speed control must consider the routes desired for each train, the moment where each route was requested and it has to verify if there are not unsafe situations in such routes. Unsafe situations are verified through the positions and speeds of other trains, switching machines position and others restrictions from possible operational commands.

2 Uses of Microprocessors in railway systems

Railway signaling exists to guaranty basic needs of safety. Its development, otherwise originates new facilities, considering the increased, performed by higher speeds and new types of control [1].

Traditional railway systems are supervised and controlled by a set of electromechanical relays that are very expensive and hard to maintain. On the other hand, they have a high reliability level. By the time, processors are replacing the electromechanical relays, offering great flexibility to maintain and upgrade the system, with higher availability.

To guarantee the correct train flow control, it is necessary to define a set of state variables logically controlled by boolean expressions. Therefore, each physic element has associated a group of boolean expressions. The origin of those boolean expressions comes from relay systems. An interesting and cheaper solution consists in import those boolean expressions and translate them to a computer code. A typical boolean expression is: $A = B * C + (.D + A)$;

The boolean expressions contains vital variables that use the “logical AND operator” and “logical OR operator”. The “logical NOT (symbolized by a dot)” inverts the value of a variable, but it cannot invert a whole expression.



In contrast with relays, which are designed to have fault tolerance, all boolean expressions in a processor are stored in memory positions, without a direct physic fault tolerance mode.

The use of processors generates new challenges in a safety analysis. Independent Safety Auditors should consider the failure mode in hardware and software in use. The safety analysis should consider alternative techniques to complete the complex task of evaluate how safe is the use of processor and its associated software [2]. Therefore, the use of simulation can improve and increase a safety analysis, searching for fault states that could not be found in a static analysis [3].

The main goal of this paper is to describe the development of a tool that simulates the behavior of trains' movement in a subway system, through the use boolean expressions. The set of boolean expressions coordinate all the movements and the simulation provides the possibility to find out lacks of safety, considering different situations and combinations in those boolean expressions. Another important goal is to simulate equipment faults in controlled conditions, in order to investigate problems not visible in a static analysis or even in a practical field test.

3 ATC environment simulation

Simulation is concerned with building a computer model of equipment or assets under study and to evaluate different scenarios, that can even explores unrealistic situations, too expensive or too harder to implement in real word [4].

This way, the use of simulation has been using to evaluate a real Railway System. The main core of this study is an ATC Environment simulation. The whole project uses the oriented-object modeling concept.

This kind of modeling provides a natural evolution in the software modules complexity. The modeling uses objects as trains, track circuits, switching machines, the ATC and others that can be interesting to use at the simulation (Figure 1).

An important ATC module is composed by one Translator (including a lexer and a parser) capable to process the boolean expression language. This first Translator is specific to the boolean expression language, but it is possible to substitute it without changing the environment simulation.

The Environment Simulator module is the object responsible to manage all other objects. It defines the time base and the state machine changes. The use of a time base represents the ATC point of view, as a discrete system, transferring, this way, snapshots from the other objects. When the ATC receives the snapshots, it has information enough to feed the boolean expressions and then update its outputs.

The Environment Simulator module has the mission to be a layer, linking the ATC and all objects. It has to control the ATC inputs and outputs, assuring a realistic approach in simulation of a railway system.



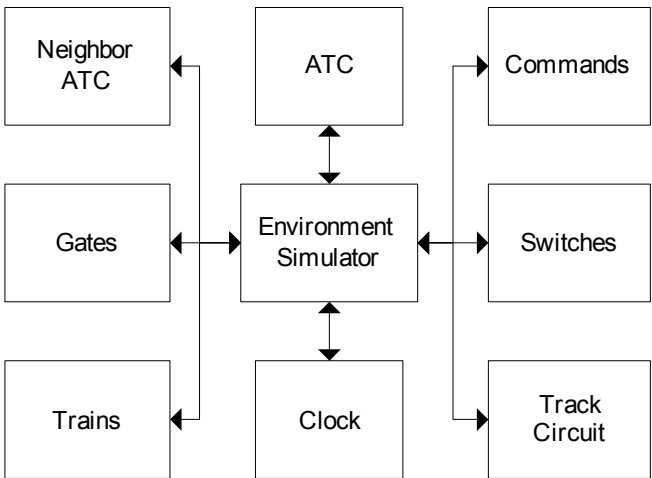


Figure 1: Example of an ATC environment modeling.

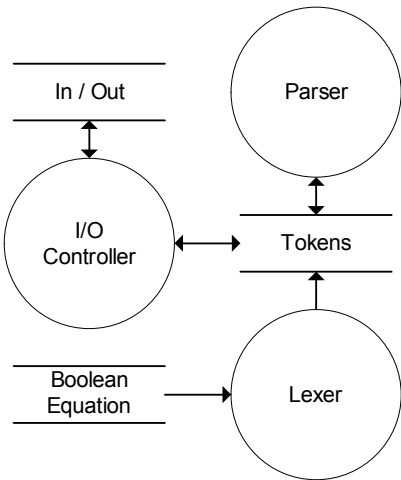


Figure 2: ATC module.

The ATC contains the Lexer, the Parser and the I/O Controller (Figure 2). The I/O controller depends on how much knowledge the designer has from a real ATC. If the ATC is a white box, the simulation will represent the I/O Controller very well. Therefore, as much detail from the real system is available as much accurate is the simulation.

After the definition of each object, using information from real systems, the Environment Simulator can overstress the ATC module. The overstress simulation offers a range of situations harder to be detected in a real railway system [5].

The first step is to execute all boolean expressions considering an initial state. The Environment Simulator controls the initial state and feed the objects with those states. Then, the simulation continues through all known relevant objects, using information about their situation. It is important to notice that the ATC can execute the boolean expressions with just the initial values of each boolean variable.

This way, even with a simple Environment Simulator, it is possible to force situations and then observe the results in a specific set of boolean expressions. That modeling offers a great range of very flexible simulation choices.

3.1 Implementation

The implementation has been developing in a Microsoft Visual Studio .NET 2003 framework. The language used is C# and all classes consider future expansion, which means, the modules can receive different configurations of railway systems domains or else different grammar rules to boolean equations. The Figure 3 shows the some classes like Parser and Lexer.



Figure 3: Example of classes used in this project.

Our implementation has an initial interface (Figure 4) which is possible to monitor all project variables. It is very important to know the value at boolean variables in each simulation step, because the Environment Simulator need those information to define the state changes of all element in simulation. The parser checks the grammar and then calculates the equations considering the entrances it receives from the Environment Simulator.

This interface offer a possibility to configure in a simple way, any variable that could concern in a simulation (one parser step) and discover which equations has updated their values. The environment uses our native language (Portuguese), but it is flexible to translate to any language.



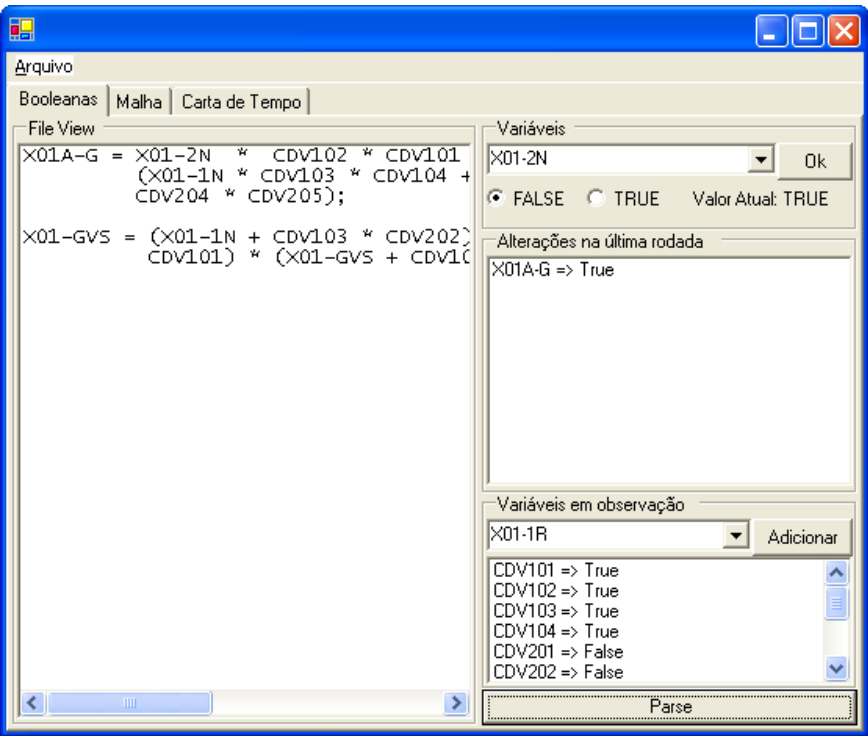


Figure 4: The parser user interface.

4 Conclusion

Preliminary results have shown that the use of a simulator to execute boolean expressions offers a great variety of tests, allowing the detection of unsafe situations, complementing software tests validation in a final release. Through simulation, it is possible to observe the behaviour of simulated objects in specific internal points. As example, such internal points can expose how much cycles the ATC executes the boolean expressions to change the value of a boolean variable.

The ATC Simulation improves a typical safety analysis, which has to deal with the completeness problem [6]. The simulation makes possible to analyse dynamic states and then offer to a safety analysis group a powerful tool to determine the reliability of a computer based railway system. The use of computer-based architecture offers a huge range of project options, instead of relays architecture. The computer flexibility can imply in unnecessary redundancies, a volume of expressions bigger than it is really necessary and feedbacks between expressions that can result in unexpected states.

In contrast with relay systems, the execution of all boolean expressions is serial. Then, the order that the expressions are written can influences the amount

of necessary cycles of update an output value. Such characteristic can hide hazard situations in a great amount of boolean expressions of a real ATC System.

Despite this environment simulator being initially focused in boolean expression, its architecture allows that the simulator stress others modules. It is possible to include more details in objects like trains, including its own controller's objects. Then, in a complete simulation, it is possible to detect how specific faults in one module can alter values in another.

Such kind of simulations has the main goal to improve aspects of safety and other RAM elements (Reliability, Availability and Maintainability).

References

- [1] Cunliffe, J.P. *A survey of railway signalling and control* Proceedings of the IEEE, Publication Date: April 1968, Volume: 56 , Issue: 4 On page(s): 653 – 674, ISSN: 0018-9219
- [2] Matsumo, M. *The revolution of train control system in Japan*, Autonomous Decentralized Systems, 2005. ISADS 2005. Proceedings, On page(s): 599 – 606
- [3] Li, J.J. *Prioritize Code for Testing to Improve Code Coverage of Complex Software* Software Reliability Engineering, 2005. ISSRE 2005. 16th IEEE International Symposium on 2005, On page(s): 75- 84, ISSN: 1071-9458, ISBN: 0-7695-2482-6
- [4] Yazdi, H. Roberts, C. Fararooy, S. *Intelligent condition monitoring of railway signalling equipment using simulation* Birmingham Univ. Condition Monitoring for Rail Transport Systems, IEE Seminar on 1998, London
- [5] M.J. Martin, J.D. *Synchronisation of cyclic coset codes for railway track circuit data* Collins, Sch. of Electron. & Electr. Eng., Bath Univ., UK; Railroad Conference, 1996., Proceedings of the 1996 ASME/IEEE Joint On page(s): 137 – 141, 1996, Oakbrook, IL
- [6] Del Frate, F., Garg P., Mathur, A., Pasquini, A., *On the correlation between code coverage and software reliability* Proc. 4th Int'l Symposium on Software Reliability Engineering (ISSRE), pp. 124-132, 1995.



This page intentionally left blank

Common approach for supervising the railway safety performance

E. M. El Koursi¹ & L. Tordai²

¹*INRETS-ESTAS, Villeneuve d'Ascq, France*

²*UIC, France*

Abstract

The key aspect of maintaining and improving safety where possible is to supervise the safety targets at European and national levels. The railway safety directive is aiming to set Common Safety targets (CSTs) that must be reached by the different parts of the rail system and the system as a whole, expressed in risk acceptance criteria. This paper intends to summarise the main results of SAMNET “Safety Management and interoperability thematic network for railway systems” project. SAMNET belongs to the Fifth Framework Programme of the European Commission and addresses the Safety Directives issues. It focuses on the development of Common Safety Targets (CSTs) that are used for supervising the safety performance. The Common Safety Indicators (CSIs) that are used to ensure the measured safety performances (e.g. the effectiveness of Safety Management System) are discussed. The Common Safety Methods (CSMs) to ensure that the targets are correctly assigned and followed at all levels (company, state, Europe) are presented.

1 Introduction

The European Community regards its railways as an economic, efficient, environmentally friendly and very safe mode of transport. The railways need a liberalised rail transport market similar to those in the civil aviation and maritime sectors. The process of creating an internal European rail market has just started. The first railway package of European Directives creates a common framework for access to railway infrastructure and for allocation of railway infrastructure capacity. The European Directives on interoperability concern the global approach to all rail transport aspects, the second railway package, published in



2004 (Directive [9] and [10]) expanded this approach by adding the safety aspects.

2 SAMNET project

SAMNET “Safety Management and interoperability thematic NETwork for railways system” (El Koursi and Tordai [7]) is a project launched by the European Commission in 2003 to investigate and to propose the approaches to specify and to implement the requirements identified in the Safety Directives. In particular, issues concerning policies on Safety Management System, Common Safety Indicators, Common Safety Targets and Common Safety Methods are addressed by this project. The main results of the SAMNET project that ended its activities on December 2005 are summarised in the document referenced SAMNET synthesis report (El Koursi et al. [5]).

3 Common tools to supervise the European railway safety

For the creation of a single European rail transport market, it is important to increase confidence between the actors on the market and between member states. For this purpose the EU Safety Directive introduces a mechanism to adopt a minimum CST (common safety targets) expressed in risk acceptance criteria for individuals and for society. *Member States shall ensure that railway safety is generally maintained and, where reasonably practicable, continuously improved, taking into consideration the development of Community legislation and technical and scientific progress and giving priority to the prevention of serious accidents.*

3.1 CST (Common Safety Targets)

According to article 7.4 of the Safety Directive (Directive [8]), CSTs shall define the safety levels that must at least be reached by different parts of the railway system (such as conventional rail system, high speed rail system, long railway tunnels or lines solely used for freight transport) and by the system as a whole in each Member State, expressed in risk acceptance criteria for:

a/ Individual risks relating to:

- Passengers
- Staff including the staff of contractors
- Level crossing users and others,
- And, without prejudice existing national and international liability rules, individual risks relating to unauthorised persons on railway premises

b/ Societal risks

The CSTs (Cassir et al. [3] and EL Koursi et al. [5]) can refer to different “groups at risk” such as passengers, staff, track workers etc. Besides these groups a distinction can be made between individual risk and collective or societal risk. Individual risk defines the chance of a person dying due to a certain activity.



Societal risk deals with the consequences of a railway accident on the environment (in terms of harm and damage).

According to Art 7.4, the Directive requires at least five different Global Safety Targets that must be reached in each Member State. These correspond to:

- Global individual risk for passengers
- Global individual risk for staff
- Global individual risk for level crossing users and others
- Global individual risk for unauthorised persons on railway premises

3.1.1 Global CSTs and Specific CSTs

Commonality for targets could therefore be considered only as applicable to objects which are to the same in all of the Member States such as generic functions, sub-systems, operations or part thereof which are characteristic of all area of the European rail system. Safety targets applied to requirements in TSIs, sub systems or Interoperability Constituents are good examples of such Specific Common Safety Targets.

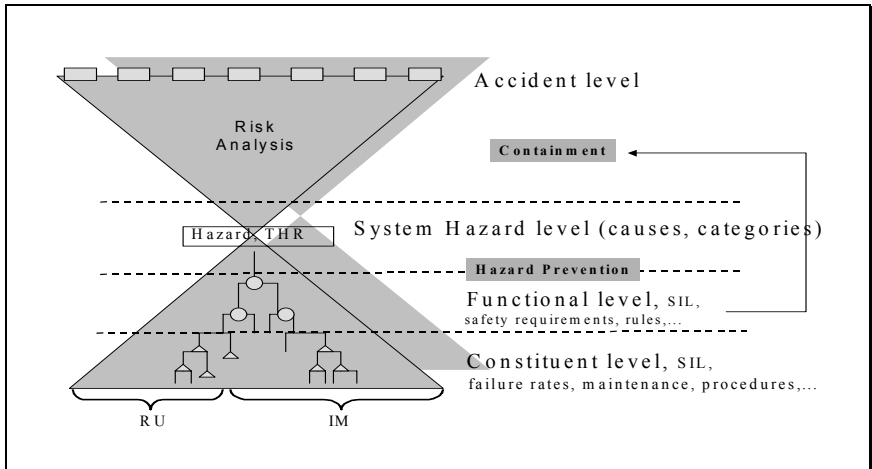


Figure 1: Proposed apportionment of safety targets.

Deriving acceptable risk levels for the various parts of the railway system requires first a classification of all the risks into various categories, and then the assignment of a target or acceptable risk level to each category with respect to each group(s) exposed to the risk. Such a process is also called risk apportionment. We can identify roughly 5 distinct approaches for the classification of risks (Mihm et al. [12]) and derivation of acceptable risk levels for parts of the railway system (Figure 1):

- System breakdown approach
- Breakdown by categories of hazard causes
- Functional breakdown approach
- Breakdown by hazard types



- Breakdown by accident types

These 5 approaches (5 possible structures for Specific CSTs) correspond in fact to different levels of detail that can be focused on when analysing safety of a railway system. This can be illustrated by the following figure showing the hierarchy in a typical safety requirement allocation process (Cassir et al. [3]).

3.1.2 Functional breakdown approach for CST apportionment

The railway system is decomposed in six sub-systems that are: **energy, infrastructure, control and signalisation, rolling stock, operation and telematic**. These subsystems are performed by a number of functions, which are implemented by a number of resources and programmes. All these elements constitute the Representative Architecture of the conventional rail system (Gigantino [11] and Chatel et al. [4]).

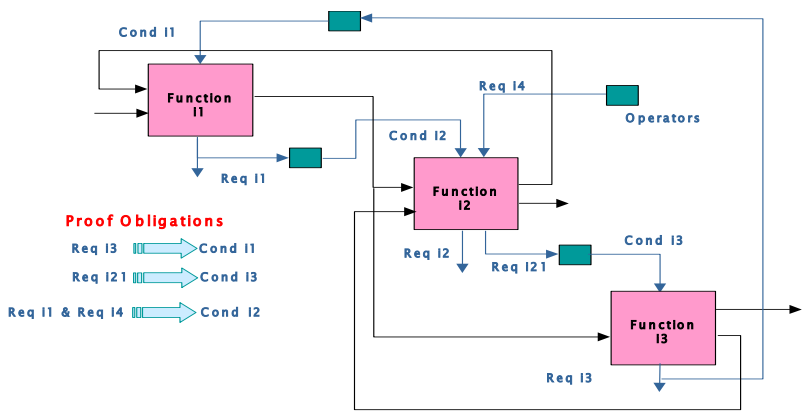


Figure 2: Proof obligations to validate the safety requirements.

This approach showed that if some functions can clearly be allocated to the IM or to the RU, responsibilities within other functions are shared. The shared responsibility aspect of some functions may derive from the goal of the function that involves both the IM and the RU or from the way it is implemented. Ensuring safe railway operations can also be expressed in a qualitative way. Both qualitative and quantitative safety targets can be allocated from a bottom-up approach as well as a top-down approach. The tool user is based on SADT technique associated with formal method based on proof obligations used to assess the allocated safety requirements (Figure 2). Global CSTs are the reference to derive CSTs for the various parts of the railway system because the specific acceptable risk levels should be derived for these different parts of the railway system as well as for the global safety targets.

We recommend the use of the AEIF [11] breakdown structure to manage the iceberg effect of setting targets at Member State level whilst we recognise that the AEIF functional breakdown is something that Member States have endorsed as a tool at Article 21 level.



3.2 CSI (Common Safety Indicators)

3.2.1 Relationship between CSTs and CSIs

As the Safety Directive is aiming to set Common Safety Targets (CSTs) for monitoring safety performance, it is essential that an agreed set of Common Safety Indicators (CSIs) be used to ensure that measured safety performance (the effectiveness of Safety Management System) of different organisations are comparable and are related to CSTs (Tordai [13]). Each of these safety targets would need to be defined with adequate Common Safety Indicators (CSIs).

Table 1: Positions on Global safety targets and Indicators.

General Common Safety Target	General Common Safety Indicator (for one Member State, per year)
Risk of death or serious injury to passengers as a result of train operations or who are harmed by any other means	Total number of deaths and of serious injuries on railway premises as a result of train operations or who are harmed by any other means (possibly scaled as “per passenger km”) - this would include falls on station platforms, accidents due to passenger behaviour etc
Risk of death or serious injury to railway employees.	Total number of deaths and of serious injuries (per track km); includes staff employed by contractors <i>It may be effective to separate track workers and on-train staffs. “Track km” may be the correct scaling parameter for track workers but it is unlikely to be correct for on-board workers</i>
Risk of death or serious injury to third parties (innocent or illegitimately present)	Total number of deaths and of injuries (per train km); includes harm to persons correctly using level crossings or living near the railway. This should also include trespassers and persons using level crossings incorrectly and contributing to the cause of the accident <i>In adopting these positions, it should be noted that in practice it is often not possible to identify innocent and culpable third parties. Care will need to be exercised by Member State to ensure a satisfactory way is developed for determining into which category each victim should be placed.</i>

This risk is expressed as a collective risk and not an individual risk, and is meant to essentially include risks of catastrophic accidents involving people and the environment, for example the release of dangerous goods. Since these events ought to be extremely rare, it is considered that a factor for scaling would not be necessary, therefore a unit per year for each country is suggested. It is suggested to use Global CSTs and Specific CSTs related to the identified parts of the railway system. One or more corresponding Common Safety Indicators (CSIs) can measure the level to be achieved by Member Sates.



General Indicators: measure the overall level of safety of the railway within a Member State (Table 1).

Specific Indicators: measure the effectiveness of the SMS of individual companies (RU & IM). The safety level of the whole system must be reached by different parts of the system. Firstly the performance of the whole system and the effectiveness of Safety Management System should be measured.

Societal risk proposed by SAMNET should contain (Tordai [13]):

- Risks that affect society as a whole, such as environmental harm. This is in addition to the total risk of harm to groups such as individual passengers, staff, level crossing users and unauthorised persons,
- Risks to persons living near the railway (railway neighbours),
- Risks of collective accidents, that is of accidents causing multiple fatalities,
- Intermodal effects, for example the fact that passengers will transfer to road because they perceive the railway to be unsafe,
- The risk that the public will lose trust in the institutions of the State.

3.3 CSM (Common Safety Methods)

According to the European Safety Directive (Directive [8]) the Common Safety Methods (CSMs) are the methods to be developed to assess whether safety targets and other safety requirements are met. Typical CSMs shall comprise methods for:

- Risk evaluation and assessment;
- Assessing conformity with requirements in safety certificates and safety authorisations (issued in accordance with the Articles of the directive), and;
- Checking that the structural subsystems of the trans-European high-speed and conventional rail systems are operated and maintained in accordance with the relevant essential requirements (as far as they are not yet covered by TSIs).

This section focuses on the CSM for Risk Evaluation and Assessment, as this is both a priority in the Safety Directive and the area where most previous activity has been focussed. A common framework for application of safety methods can consist of:

- System Definition
- Identification of Hazards
- Risk Analysis
- Risk Evaluation
- Risk Reduction
- Risk Control
- Risk Monitoring
- Risk evaluation and assessment

The urgent priority in the Safety Directive (Directive [8]) is for the development of a CSM to address risk evaluation and assessment. A literal interpretation of the Safety Directive implies that the CSM for risk evaluation and assessment



includes just a method for assessing the risk associated with the railway system, and comparing those against a set of defined risk criteria (i.e. System Definition, Identification of Hazards, Risk Analysis and Risk Evaluation). ‘Risk Evaluation and Assessment’ must support the development of an organisation’s safety requirements and the procedures that comprise the SMS. The generic CSM framework for assessing conformity with the typical elements that may be required when safety certificates and safety authorisations are issued in accordance with Articles 10 and 11 of the directive, consist of, amongst others, Risk Reduction, Risk Control and Risk Monitoring.

3.3.1 CSMs for risk evaluation and assessment

There is an inherent difficulty in allowing a variety of tools and techniques to be applied for risk analysis and assessment (Bearfield et al. [1]). It is therefore important to establish the degree of commonality that is actually required between member states in the application of the CSMs. There are three key principles:

- Commonality of the System Definition. The System Definition activity provides the scope of all subsequent risk identification, analysis and management activity. The requirement should be that any systems breakdown is consistent with the interoperability directives Product Breakdown Structure or the AEIF Functional Breakdown Structure (at least at the highest level of abstraction).
- Commonality of Conceptual Model used to undertake Risk Assessment. The bow-tie concept (Figure 3) is the generally accepted conceptual model used to structure risk analysis and assessment. There are various tools and techniques that can be used to elaborate this conceptual model and undertake more detailed risk analysis depending on the depth of analysis required and the nature of the accident and its various causes.
- Commonality of Base Units of Risk. ‘Risk’ is accepted to be the product of the probability of occurrence of an accident and the severity of that accident. However there are various and different approaches adopted across each member state for quantifying risk and a number of units of risk used. If the base units for risk are standardised then the results of any analysis should ultimately be comparable.

Initial use of the techniques indicates that the “Eurosig” scenarios provide a good basis for analysis.

The ‘bow-tie’ model is not a tool or technique for risk analysis. It represents the underlying mental model used by most safety engineers to undertake risk analysis. The majority of risk analysis tools and techniques develop and formalise this representation in various ways.

3.1.1.1 Risk evaluation Once the risks associated with the various hazards have been calculated it is necessary to evaluate their acceptability. The tolerability of calculated risks should be evaluated against the CSTs derived for the project. *The total number of deaths and injuries should be weighted against some agreed criteria such as ‘equivalent fatalities’. The ‘equivalent fatalities’*



would represent an aggregation of ‘fatalities’ and ‘major’ injuries according to a ratio yet to be decided (e.g. 1 fatality = 10 major injuries = 1 equivalent fatality). The current proposal is that minor injuries would not be included in the measures, as they tend to indicate risk associated with low severity accidents.

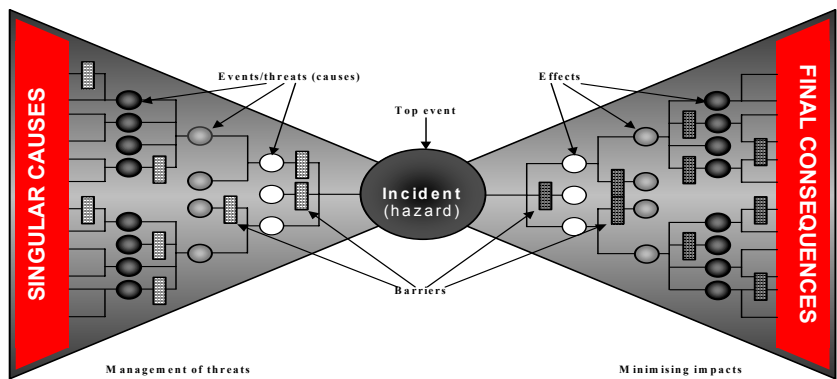


Figure 3: Bow-tie model.

3.3.2 CSMs for meeting safety requirements

The CSMs for meeting safety requirements (Directive [8]) are related to:

- Defined in the safety certificate of each RU
- Defined in the safety authorisation for each IM
- Derived for structural sub-systems not completely covered by TSIs

CSMs are required for ensuring that the various safety requirements put in place by an organisation are routinely and continually met by the organisation. This CSM should comprise of the risk reduction, risk control and risk monitoring stages of the risk assessment process.

3.3.2.1 Risk reduction The possible approaches to risk reduction are:

- Revision of the technical system design
- Modification of operational procedures
- Changes to staffing arrangements
- Training of personnel to cope with the hazard

These risk reduction approaches essentially describe revisions that an organisation can easily make to its safety requirements.

3.3.2.2 Risk Control and Monitoring Risk Control and Monitoring is about how an organisation understands the practical effectiveness of the safety requirements. The organisation must monitor all of its barriers to ensure that failures or weaknesses in the barriers are identified and rectified before an accident actually occurs. The barriers in an organisation are its safety requirements. Each different type of barrier should be monitored using a different process. For example, where the safety requirements are operational



procedures they can be controlled and monitored using safety inspections, and auditing. Where the safety requirements are functional system requirements they can be monitored by inspection or they may also have diagnostic functions which aid in understanding where safety requirements are no longer being met or are ineffective. A common tool to assist risk reduction and control is the use of a Hazard Log. This involves recording all identified hazards and actions that mitigate risk in the Hazard Log.

4 Conclusion

The basic element to develop the common safety targets and methods is the establishment and the agreement on common definition of railway system. The approach proposed in this paper allows the definition of global safety targets according to the safety directive and the development of specific targets related to specific parts of the system based on the functional architecture of railway system developed by AEIF "Association Européenne pour l'Interopérabilité Ferroviaire". The purpose of Common Safety Indicators "CSIs" is to monitor European railway safety performance. It is essential that an agreed set of common safety indicators should be used to ensure that measured safety performance of different organisations are comparable and are related to Common Safety Targets "CSTs". In this paper, we suggest the definition of global safety indicators in relation with the global safety targets prescribed by the safety directive. The specific indicators can be used for critical events and not be based on accidents (this is covered by Global Indicators). The common safety methods can be based on existing tools and techniques used to develop a safe railway system. The Common Safety Methods "CSMs" are required for ensuring that the various safety requirements put in place by an organisation are routinely and continually met by the organisation. The CSM process should be based on existing known techniques (e.g. CENELEC 50126).

References

- [1] Bearfield, G et al (2005), D1.3.2: "Common Safety Methods Position Paper", www.samnet.info, 11/07/2005.
- [2] Bearfield, G, S. Mitra and E.M. El Koursi (2004). Guidelines for Safety Management System, D2.2.2/V3.0, www.samnet.info, May 2004.
- [3] Cassir, C et al (2005), D1.3.3: "Specific CSTs report Case study", www.samnet.info, 10/06/2005.
- [4] Chatel, V, E.M. El Koursi, C. Felliott, U. Huisman (2002) "Functional analysis of the sub-system of energy and infrastructure of conventional rail", SMC, 2002, Hammamet, Tunisie.
- [5] El Koursi E.M, S. Fletcher, L. Tordai, J. Rodriguez (2006-a) "SAMNET synthesis report – safety and interoperability, www.samnet.info, February 15, 2006.



- [6] El Koursi, E.M, N. Duquenne (2006-b), "Assessment of the Safety management system in railway sectors", WCRR2006, Montréal, Canada, 4-8 June 2006.
- [7] El Koursi E.M. and L. Tordai (2003), "SAMNET: Safety management and interoperability thematic network in railway systems", WCRR2003, Edinburgh pp: 198-202, (2003).
- [8] Directive 2004/49/EC (2004) on safety on the Community's railways and amending Council Directive 95/18/EC on the licensing of railways undertakings and Directive 2001/14/EC on the allocation of railway infrastructure capacity and the levying of charges for the use of railway infrastructure and safety certification (Railway Safety Directive), L164 p44-113, 29 April 2004, Official Journal of the European Union.
- [9] Directive 2001/16/EC (2001) of the European Parliament and of the Council on the interoperability of the trans-European conventional rail system, Commission of the European Communities, Brussels, 19 March 2001.
- [10] Directive 96/48/EC (1996) on the interoperability of the trans-European high speed rail system, Commission of the European Communities, Brussels, 23 July 1996.
- [11] Gigantino, A (2002) 'Report on the Representative Architecture', AEIF, 19 September 2002.
- [12] Mihm, P et al (2004), D2.4.0: "Acceptable Risk Level", www.samnet.info, June 2004.
- [13] Tordai, L "Common Safety Indicators and Common Safety Targets of European Railways", WCRR2006, Montréal, Canada, 4-8 June 2006.



Potential dangerous object detection on railway ballast using digital image processing

P. L. Mazzeo, E. Stella, M. Nitti & A. Distanto

*Istituto di studi sui sistemi intelligenti per l'automazione,
Consiglio Nazionale delle Ricerche, Italy*

Abstract

The correct assessment of the condition of a railroad requires the consideration of different factors. Some factors, such as the condition of the ties, can be measured by inspecting features visible from the surface of the railway. Other factors include the condition of the ballast; it is important to recognize the critical situation in which any foreign object can be present on the ballast. These kinds of objects could be cans, pieces of sheet and everything over a well determined dimension. Extensive human resources are currently applied to the problem of evaluating railroad health. The proposed visual inspection system uses images acquired from a digital line scan camera installed under a train.

Here we focus on the problem of foreign object detection in the railway maintenance context. To obtain this aim we train a Multilayer Perceptron Network (MLPN) with the edge histogram of the ballast patches manually extracted from the acquired digital image sequence. The general performances of the system, in terms of speed and detection rate, are mainly influenced by the adopted features for representing images and by their number. By this inspection system it is possible to aid the personnel in railway safety issues because a high detection rate percentage has been obtained. We show the adopted techniques by using images acquired in real experimental conditions.

Keywords: *obstacle detection, ballast inspection, neural networks.*

1 Introduction

Inspection of the rail state is one of the basic tasks in railway maintenance. In the last few years a large number of methods have been proposed by the computer vision community for facing the problem of visual inspection [1, 2]. The



problem is a particular instance of the most general question of detecting objects in digital images as faces [3], pedestrian [4], balls [5], just for citing a few examples. Recently, such methods have been successfully applied for railway inspection and monitoring [6, 7]. In fact, the growing of the high-speed traffic on the rail tracks demands the development of sophisticated real-time visual inspection systems. The railroad network must be continuously maintained to assure the safe and timely delivery of freight and passengers. This maintenance effort is complicated by the fact that accurate assessments of track condition are difficult and expensive to obtain. In many cases, efficient data collection tools exist, but human experts are required to analyze the data and identify sections of track that require maintenance. With thousands of miles of track, this task is very often impractical. In this paper, we apply pattern recognition techniques in order to automate the analysis of railroad data. Normally the maintenance of the railway plane is done by trained personnel who periodically observe the images recorded by a TV camera installed on a diagnostic coach. This manual inspection is lengthy, laborious and potentially hazardous and the results are strictly dependent on the capability of the observer to catch possible anomalies and recognize critical situations.

A considerable body of work has been published on the problem of applying recent technology advances to aid in the classification and management of railroad systems. Computers have been used to help automate the collection and analysis of railroad data as early as 1967 [9]. Sensing systems capable of collecting various track geometry data have been developed and applied. Systems capable of recording track geometry and and strength have been developed for railroad cars [10] and high-rail vehicles [11]. Computer vision systems have been applied to monitor rail wear [12], measure rail width [13], and produce a database of continually scanned visual images [14]. These systems facilitate the automated collection of large amounts of railroad data. However, the data must be parsed and analyzed by a human expert before it can be used to make maintenance decisions.

Similar advances have been made in the context of detecting of sleepers' anomalies, as well as missing fastening elements, is an important task that an efficient inspection tool should supply. As described in our previous work [6] the Wavelet Transform has been successfully applied in railway context for the recognition of fastening elements. In other works [8, 15] this fastening element are recognized by using Independent Component Analysis (ICA) and Support Vector Machine (SVM). This kind of detection problem can be regarded as the problem of detecting flat objects from 2-D intensity image. Usually, such problems have been approached by using algorithms of edge detection, border following, thinning, straight line extraction, active contour (snake) following [16]. In this work we focalize on the ballast recognition in order to catch the possible presence of every potential dangerous foreign object (cans, piece of sheet etc. plastic bottles) that do not belong to railroad structure.



There are few works that focus on the measurement of ballast and subgrade condition. In these works the primary tool for non-destructive measurement of subsurface condition is ground penetrating radar (GPR). GPR has been used to identify fouled ballast and trapped water on an experimental railroad test bed [17]. In [18] GPR was found particularly useful for locating drainage problems and identifying areas of similar substructure condition. Other efforts have been successfully combined GPR with infrared imaging systems to detect subsurface defects in railroad track beds [19]. However no works have been found on ballast inspection in order to recognize potential dangerous object that not belong to railroad bed. The main idea of this work is to find a non-invasive technique that use the images captured by a TV Camera. To reach this aim it has been found a technique based on the edge-histogram combined with a neural classifier (Multi Layer Perceptron). The main idea behind this technique is that any foreign object that does not belong to railroad ballast has a well-defined structure. In this way we can characterize two object classes: the ballast, and anything that is not ballast. The texture classification techniques are not used because the ballast is not a texture. It is not a function of the spatial variation in pixel intensities (gray values) which can characterize the ballast distribution. Furthermore, the ballast has very different pixel intensity (a great variability of gray level luminosity) and the dimension of each stone is strongly variable. In this work we present a collection of cases in which our technique works as ballast detector and we present a table where a detection rate percentage gives a measure of the robustness and reliability. This paper is organized as follows: In section 2 an overview of the implemented system is given. In section 3 we briefly describe of the pre-processing used technique. The MLP classifier is succinctly introduced in section 4. Finally experimental results are given in section 5.

2 System overview

As introduced in the previous section, we have realized a vision architecture that is able to detect the ballast region in the digital image acquired. This system identifying the ballast region can account the critical situation in which any potential dangerous object is presented on the ballast surface. This type of anomaly is very dangerous because when it occurs can obstacle the normal train traffic, generating consequences for safety issues. In truth the potential dangerous objects can lie on the sleepers but in this case they will be treated in different way.

The proposed system architecture is based on data classification by using a supervised learning scheme. The logical scheme of the whole system's processing block is showed in figures 1 and 2. The system is composed by 3 main blocks:

- Prediction Algorithm Block (PAB);
- Rail Detection and Tracking Block (RD & TB);
- Ballast Identification Block (BIB).



PAB extracts from an input video sequence the images areas candidate to contain the sleepers. This block takes advantage of the distance between two consecutive sleepers that is a priori known distance.

Rail Detection and Tracking Block determines the rail position respect than acquired images and tracks it in the video sequence. BIB examines the areas where is the ballast and determines the situations in which potential dangerous objects are presented.

By this blocks the systems is able to identify the sleepers (PAB) and it enables the Ballast Identification Block (BIB) only in that areas where is present the ballast. Observes that without the PAB the system could classify the sleeper regions as foreign objects because it does not belong to the ballast class. This algorithm jointly to Rail Detection and Tracking Block (RD & TB) works to determine the region candidate to contain the ballast. PAB [20] uses the rail position (determined by RD & TB) for calculating the distance between two consecutive sleepers in order to predict the position of the image areas candidate to contain the ballast patches. In figure Y is exploded the Ballast Identification Block (BIB) and all its functional parts are shown.

This Block is subdivided in two functional sub-blocks:

- Edge Detection & Histogram Block (ED & HB);
- Multi Layer Perceptron Network Block (MLPNB).

The first block performances a pre-processing on the input patches in order to extract the features which characterize the images. The extracted features are the input coefficients of the MLPN contained in the second block.

The reduction of the input space dimension performed by the ED & HB is the first step to successfully speed up the whole classification process. This target can be reached by using a feature extraction algorithm which should be able to store all the significant information about the input patterns in a set with few important elements. To achieve these requirements we have used the edge-histogram pre-processing technique as shown in figure 2. Used approach allowing to capture the object structure that distinguish the potential dangerous object between the ballast.

The output of ED & HB is a coefficient features vector that is the input of a MLPNB. More details are equipped in the next sections. Finally BIB produces a Pass/Alarm signal which is online displayed. In case of alarm (i.e. potential dangerous object detected on the ballast) BIB records the coordinate into a log file.

3 Pre-processing

As delineated in previous section the implemented system extracts from the candidate patches the edge histogram. This method captures non-directional edges as well as any directional ones. For the edge extraction step we have used the Sobel edge detector mask in the horizontal and vertical direction.



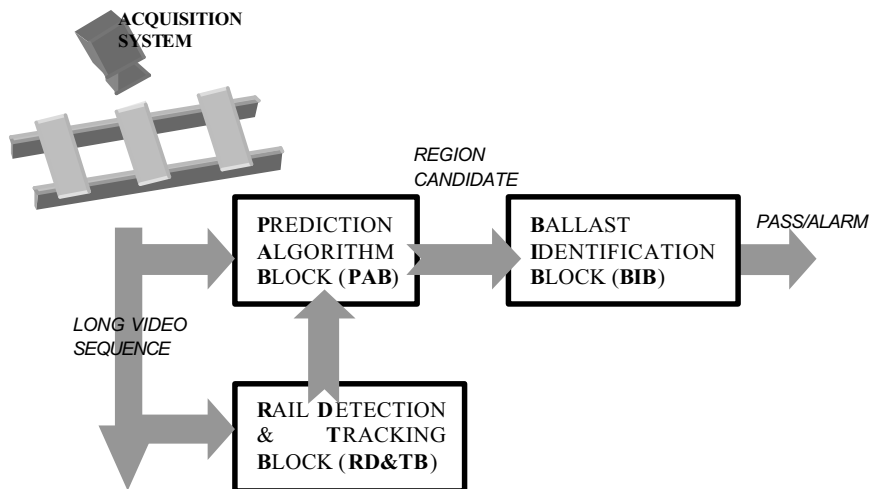


Figure 1: System architecture.

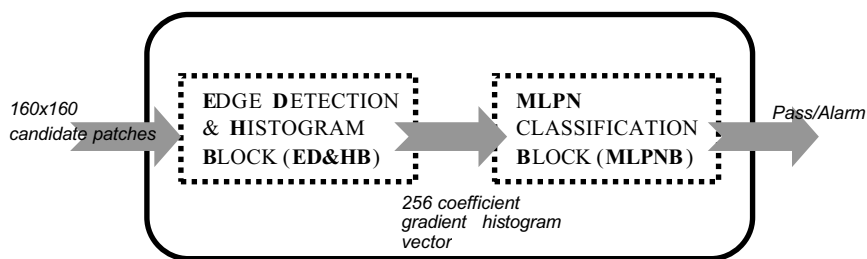


Figure 2: Ballast identification block.

After that we have employed these edge image results to find the approximate absolute gradient magnitude at each point in an input grayscale image:

$$|G|=|G_x|+|G_y| \tag{1}$$

where G_x estimating the gradient in the x-direction (columns) and the G_y estimating the gradient in the y-direction (rows).

Finally we calculate the edge histogram on the gradient image G , based on 256 bins that is the number of the grey level images.

Figures 4 and 5 represent the edge histogram of the two image patches that are shown in figure 3.





Figure 3: Examples of 160x160 pixels patches: foreign object (left) and ballast (right).

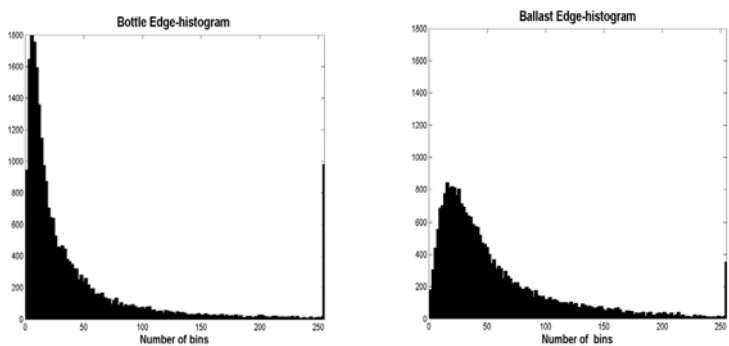


Figure 4: Ballast edge-histogram (right side), dangerous object edge-histogram (left side).

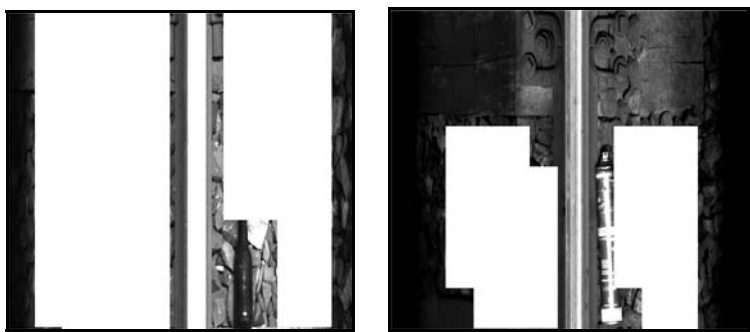


Figure 5: Examples of the system output image.

4 Classification

The classification process has been implemented by using artificial neural networks. Neural computing is particularly successful at extracting data from images. In our particular context of object detection neural-based methods have a key advantage over geometry-based methods because they do not require a geometric model for the object representation. In our detecting system a Multilayer Perceptron trained by Back-Propagation algorithm has been used (MLPN). This architecture consists of three-layers of neurons (input, hidden and output layer). The input layer has a number of neurons equal to the number of image features extracted in the pre-processing step.

The output layer has only one unit which generates an output ranging from 0 to 1 indicating a measure of confidence on the presence of the object to detect in the current image patch. As activation function for both hidden and output units we have used the log-sigmoidal function that generated output between 0 and 1. Error Back Propagation algorithm with an adaptive learning rate has been used to solve the weights of the MLPN [21].

5 Experimental results

The images of the rail have been obtained by a line scan camera DALSA PIRANHA 2 with 1024 pixels of resolution (maximum line rate of 67 kLine/s) with the transmission protocol Cameralink, installed under a diagnostic train during its maintenance route. Furthermore we have used the frame grabber PC-CAMLINK (Imaging Technology CORECO). In order to reduce the effects of variable natural lighting conditions, an appropriate illumination setup equipped with six OSRAM 41850 FL light sources has been installed too. In this way the system should be robust against changes in the natural illumination. Moreover, in order to synchronize data acquisition, a trigger is sent to the TV camera by the wheel encoder.

The spatial resolution of the trigger is 3 mm. A pixel resolution of $1 \times 1 \text{ mm}^2$ can be obtained choosing a TV camera with focal length of 12 mm. The integration time of the TV camera has been properly set in order to acquire images at maximum speed of 241 km/h choosing the spatial resolution of 1 mm. A long video sequence of a rail network has been acquired in order to experiment the proposed visual-based inspection system. Firstly, a number of sample images has been extracted from the sequence to create the training set for the neural classifiers. The remaining video sequence has been used to test the performance in term of detection rate e computational velocity of the developed inspection system.

Positive and negative examples of the ballast patches have been manually extracted from the sequence training images (see figure 3 right side). In the experiment stage we have used a crescent patch dimension in order to determine the best compromise between the ballast identification process and the consequent potential dangerous objects identification. Anyway maximum



dimension of the example patches consist of a 160x160 pixels subwindow (see figure 3).

This training set contains 520 positive examples (ballast) and 183 negative examples (foreign objects). The used Multi Layer Perceptron is composed by 256 input layer neurons 30 hidden layer ones and one output neuron. The same neural classifier has been trained by the edge histogram as explained in section 3 in this way the input space dimension has been reduced to 256 feature coefficients. The neural classifiers have been trained on the same number examples training sets. In order to evaluate the generalization ability of the classifier a preliminary test has been carried out on a validation set.

This set contains 800 positive examples (ballast patches) and 300 negative ones (foreign object patches). In table 1 the results of that test are shown. In the first column of Table 1 the used patch dimension are listed. Detection rates are given in terms of true positive (TP) and true negative (TN). From this table can be found out that the different trained classifiers perform almost well in all the considered cases of the patches dimensions. However the best result obtained corresponds to the last row in the table 1.

Table 1: System detection rate on a validation set.

Patch dimensions	True Positive (%)	True Negative (%)
40x40	87.98	94.55
80x80	91.92	93.95
160x160	95.38	97.56

For this reason, the patch dimension of 160x160 pixels have been chosen for the second phase of testing experiment. In that experimental phase a long video sequence of rail has been inspected by the implemented vision system. The system inspects only the image areas that are candidates to contain the ballast.

Figure 5 shows how the system works. The white regions are the patches that system identifies like ballast and the reader could be see that the foreign object are correctly identified.

6 Conclusion and future work

This paper has proposed a prototype able to detect autonomously potential dangerous object on the ballast region. The periodical inspection of railway infrastructures is very important to prevent dangerous situations. The developed system can be used to detect probable dangerous object on the ballast region in railway infrastructure.

The inspection system uses images acquired by a digital line scan camera installed under the train. Neural classifier has been trained to recognize ballast patches in railway track. Firstly the images are pre-processed by using edge-

histogram method. Then the obtained detecting system has been tested on a validation set to establish the best patch dimension which give us the best performance. After that, a long sequence of tests on real images has been examined. The system correctly detects the ballast region as well as the presence of foreign object. By our inspection system is possible to aid the personnel in the railway safety issue because high percentages of detection rate has been obtained showing a high reliability and robustness.

References

- [1] O. Silven, M. Niskanen, H. Kauppinen. Wood inspection with non-supervised clustering. *Machine Vision and Applications* (2003) Vol. 13 No.5-6: 275-285.
- [2] Y. Zhang, Z. Zhang, J. Zhang. Deformation visual of industrial parts with image sequence. *Machine Vision and Applications* (2004) Vol. 15 No.3 115-120.
- [3] K. Sung and T. Poggio. Example-based Learning for View-based Human Face Detection. *Artificial Intelligence Laboratory*. Massachusetts Institute of Technology, Cambridge, MA. 1994. A.I. Memo No. 1521.
- [4] C. Papageorgiou and T. Evgeniou and T. Poggio. A trainable pedestrian detection system. *Proceedings of Intelligent Vehicles*. Stuttgart, Germany, October, 1998.
- [5] N. Ancona and G. Cicirelli and E. Stella and A. Distanto. Ball detection in static images with Support Vector Machines for classification. *Image and Vision Computing*. Elsevier 2003. 21(8) 675-692.
- [6] P.L. Mazzeo, M. Nitti, E. Stella and A. Distanto, "Visual recognition of fastening bolts for railroad maintenance" , in *Pattern Recognition Letters*, vol. 25 no. 6, pp. 669-677 2004.
- [7] M. Nitti, P.L. Mazzeo, E. Stella and A. Distanto "Real time Position Estimation of overhead line in railway application" in *Computers in Railway IX*, WIT Press, pp. 835-843, 2004.
- [8] Mazzeo P.L., Ancona N., Stella E. and Distanto A.: Visual Recognition of hexagonal headed bolts by comparing ICA to Wavelets. *Proceedings of the IEEE Inter. Symposium on Intelligent. Control-Houston*(2003) 636--641.
- [9] Mechanized automed, computerized, *Railway and Track Structures*, March (1971) 18—21.
- [10] W. Ebersohn and E.T. Seling, Use of geometry measurement for maintenance planning. Transportation Research Record No. 1470, *Railroad Research Issues* (1994) 84—92.
- [11] A.M. Zarembski and W.T. MacCarthy, Development of nonconventional tie and track structure inspection systems. Transportation Research Record No. 1489, *Railroad Trasportation Research* (1994) 26--32.
- [12] D.L. Magnus, Non contact technology for track speed rail measurement (ORIAN), *Proc. of the SPIE Nondestructive Evaluation of Aging Railroads*, 2458 (1995) 45—51.



- [13] R.B. Ryabichenko, S.B. Popov, and O.S. Smoleva, CCD photonic systems for rail width measurement, *Proc. of the SPIE Photonics for Transportation*, 3901 (1999) 37--44.
- [14] H. Sasama, Maintenance of railway facilities by continuously scanned image inspection, *Japanese Railway Engineering*, 27 (1994) 1--5.
- [15] Mazzeo P.L., Ancona N., Stella E. and Distante A.: Fastening Bolts recognition in Railway images by Independent Component Analysis. *Proc. 3rd Visualization, Imaging and Image Processing-Benalmadena*(2003) 668—673.
- [16] Andrade-Cetto J. and Kak Avinash C.: Object Recognition. *Wiley Encyclopedia of Electrical Engineering*. vol. Sup. 1 (2000) 449—470.
- [17] G. Gallagher, Q. Leiper, M. Clark and M. Forde, Ballast evaluation using ground penetrating radar, *Railway Gazette International*, February (2000), 101--102.
- [18] T.R. Sussmann, F.J. Heyns and E.T. Seling, Characterization of track substructure performance, Recent Advances in the Characterization of Pavement Geomaterials, *Geotechnical Special Publication, American Society of Civil Engineers* (1999), 37—48.
- [19] G.J. Weil, Non-destructive, remote sensing technologies for locating subsurface anomalies on railroad track beds, in *Proc. of the SPIE, Nondestructive Evaluation of Aging Railroads* 2458(1995), 74--81.
- [20] P.L. Mazzeo, E. Stella, M. Nitti, A. Distante, Visual Recognition of fastening bolt in railway maintenance context by using wavelet transform, *ICGST International Journal on Graphics Vision and Image Processing Special Issue on Wavelets*, May 2005, 27- -34.
- [21] M. Bishop, *Neural Networks for Pattern Recognition* , New York 1999, 220 – 313.



Thermal characteristics of novel brake friction materials for light rail transit applications

N. Valliyappan, D. Berhan, M. N. Darius & G. Solomon

Faculty of Mechanical Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia

Abstract

A batch of five formulations (SP1, SP2, SP3, SP4 and SP5) of novel brake friction materials are being developed and tested for light rail transit (LRT) applications. This paper presents the thermogravimetric (TG) and derivative thermogravimetric (DTG) analysis coupled with the mass spectroscopy (MS) of those friction material samples. The LRT braking system operates at elevated temperatures and may reach an upper extreme of 900°C. In this research, temperature range from ambient to 600°C is considered the temperature zone that signifies the limits of normal braking operation and considerable weight loss. In general, all samples except SP4, decompose in four steps and registered significant weight loss averaging 5.53% ($\pm 0.20\%$) of the original weight. Sample SP4 experienced the highest weight loss of 5.77% in three steps. Samples SP1 and SP5 were found to be stable, temperature wise, in decomposing compared to all other samples. Sample SP1 exhibited a controlled rate of decomposition apparently over a wider temperature range (230°C – 688°C) with a total loss of weight of 5.37%. In contrast, sample SP5 decomposed in a greater magnitude over a narrower temperature gap (303°C – 483°C) albeit the total weight reduction is 5.30%. MS signals indicating mass number of 2 and 44 was detected in all samples at critical temperature zones corresponding to substantial weight loss. This denotes the probable evolution of compounds consisting of hydrogen, helium, carbon dioxide, nitrogen oxide and propane and is related to the physical-chemical reaction which is explained in this paper.

Keywords: thermogravimetric analysis, mass spectroscopy, friction material, light rail transit.



1 Introduction

Friction materials are complex composites comprising steel and ceramic fibres, abrasives, lubricants, metallic and semi-metallic fillers [1]. Friction materials, also generally known as brake pad, should be able to sustain thermal shocks and surface cracks besides other considerations such vibration damping and noise suppression. The friction and wear performance of friction materials is a function of the material composition, microchemical structure of the pad, rotating speed, pressure and contact surface temperature [2–4].

Brake pads are exposed to large thermal stresses during routine braking and may be subjected to unusual thermal load cycles during hard braking. High-*g* decelerations of passenger trains are known to generate temperatures as high as 900°C in a fraction of a second [5]. Brake operating conditions related to heat generation include braking pressure and rubbing velocity. However, in normal operating conditions, braking should cause a temperature increase up to around 500°C [6]. Mutlu et al. [7] studied about the tribological changes as a function of temperature up to 400°C. It is therefore expected that the generation of heat during braking and hence the elevation of temperature could bring about chemical reactions which tend to cause decomposition of constituent materials. Consequently, the chemical constitution of the brake pad continuously varies from its original composition, releasing gases in the process of thermal degradation.

The method of detecting changes in weight as a function of temperature under defined atmospheric conditions is termed thermogravimetric analysis (TG). Derivative thermogravimetric (DTG) is the time rate of reaction derived from the TG results. In view of the possibility of materials discharging gases during the temperature ramp, TGA is usually approached in tandem with evolved gas analysis (EGA) such as mass spectroscopy (MS). The combination of thermal and evolved gas analyses is highly recommended and used extensively in a variety of materials research initiatives [8–10].

This paper presents the TG/DTG-MS coupled analysis of five novel formulations of brake friction materials, which are intended for final usage in the LRT braking system in Malaysia. Characteristic temperatures corresponding to initial weight decomposition and maximum rate of decomposition are highlighted and related to the possible molecule release at those temperatures.

2 Experimental

Mettler Toledo TGA/SDTA851e was used for the TG-MS coupled analysis. The mass spectrum of the gaseous products which evolved during heating is obtained by coupling the TG thermobalance on-line to the MS via a heated quartz glass capillary tube. Samples were heated from 30°C to 1000°C at a scan rate of 10°C/min under flow of 50 ml/min of nitrogen. The TG/DTG-MS spectra were generated by plotting the continuous change of sample weight and ion intensities (mass-to-charge ratio) of the gaseous products as a function of temperature.



3 Results and discussion

3.1 TG/DTG of raw materials and commercial brake pad

There are more than 25 individual compounds in form of fibres and solids used for the preparation of the five new brake pad formulations. This section provides a brief overview of the TG/DTG results for the some of the raw ingredients namely, barium (filler), rubber (organic additive), Nipol (a proprietary polymer-based material) and iron oxide (modifier), which are used to explain the TG-MS results of the samples.

The TG/DTG curves for the commercial brake pad are shown in fig. 1. The material decomposed in two stages beyond 200°C. The polymeric binder decomposed up to 550°C at a rate of below 0.5% per minute. Total weight loss is 4.79% up to 680°C. This marks the reference amount of total weight loss for the five new brake pad formulations.

Fig. 2 shows the TG/DTG thermogram of barium. The sample gained weight to a maximum of 100.07% at rate of 0.25% at 180°C, which could be due to reaction with the surrounding atmosphere. After 380°C, weight loss was registered in three stages with maximum loss at a rate of 0.15% at 546.3°C.

The thermal decomposition pattern for rubber is given by fig. 3. This sample was stable under heating up to 245°C but decomposes in one-step until the 550°C after which no weight loss was registered.

Fig. 4 shows the TG/DTG results for Nipol (a proprietary polymer-based material) indicating a 3-stage decomposition as the sample is heated. The significant loss of weight is at the second stage for which the temperature bandwidth is between 355°C and 535°C. This sample can decompose at a rate as high as 32% per minute, hence its chemical reaction to high temperature.

The TG/DTG plot for iron oxide is shown in fig. 5. It is observed that at 49.78°C the sample gains weight up to 0.05%, possibly because of further oxidation during initial heating. Similar result was analysed elsewhere [11]. The two-stage decomposition process begins at around 220°C. The first stage occurred between 221.2°C and 345.6°C. Above this temperature, the sample underwent gradual increment of weight loss (in %; DTG) up to 0.24% at 996.3°C. The curve indicates infinitesimal decomposition slightly above 700°C.

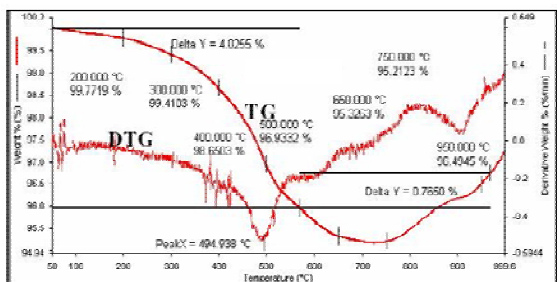


Figure 1: TG/DTG curve of commercial brake pad.



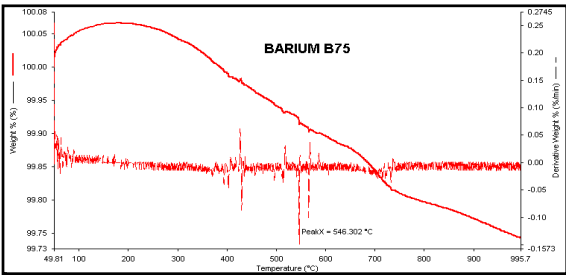


Figure 2: TG/DTG curve of barium.

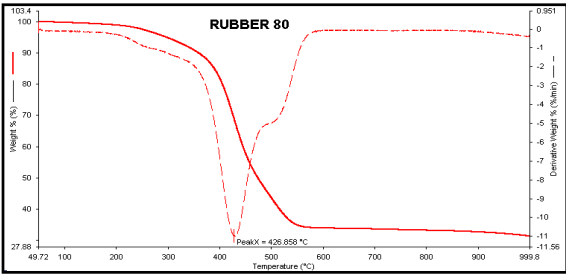


Figure 3: TG/DTG curve of rubber.

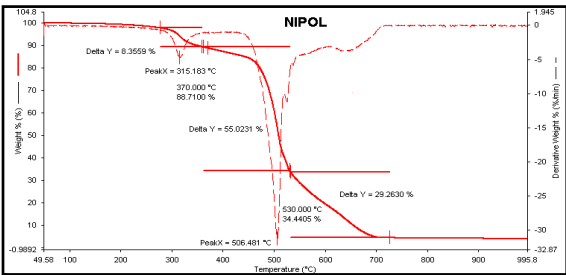


Figure 4: TG/DTG curve of Nipol.

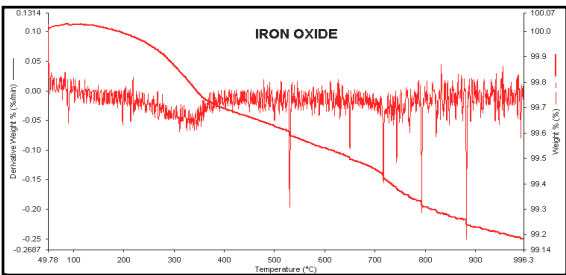
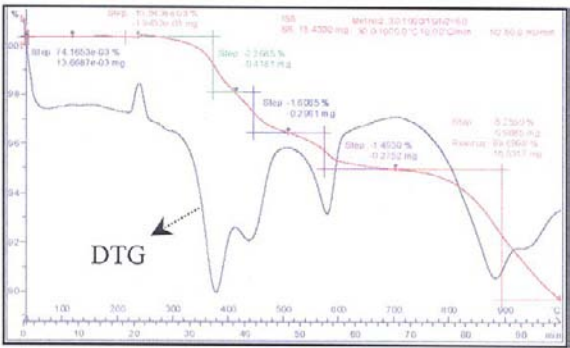


Figure 5: TG/DTG curve of iron oxide.

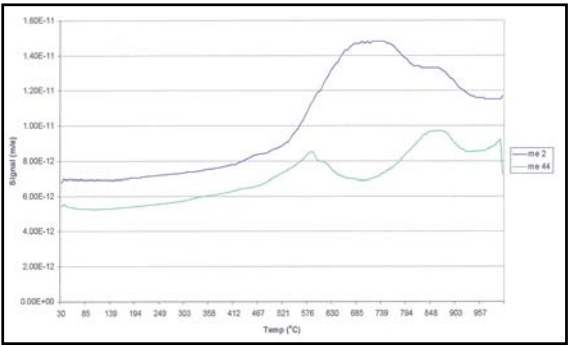
3.2 TG/DTG-MS of brake pad samples

This section elaborates the TG/DTG-MS results obtained for the five novel brake pad formulations developed by our research group. Discussion of the thermal characteristics of the samples will be based on the results of the five raw materials provided in the preceding section besides the EGA results which evidence the decomposition of the samples with increase in temperature. The temperature range of interest is up to 600°C for which calculation of total weight loss and thermal degradation pattern is accounted for in the discussion.

Fig. 6(a) shows the TG/DTG curve of sample SP1 while fig. 6(b) is the mass spectrum for the EGA. This sample decomposes in four steps. The initial weight loss occurred between before 115°C. Weight loss of 2.27% was registered at the interval of 230°C to 403°C (third step) and minimum DTG temperature, $T_{DTG, min}$, of 363°C but no MS signal was recorded. This suggests that no gaseous product(s) was released in this interval hence possibility of evaporation of the organic compounds is excluded. However, MS signals (refer fig. 6(b) and table 1) were detected in the 4th and 5th weight loss regions (<600°C) due to probable release of organic and inorganic gases as may be caused by removal of carbon.



(a)



(b)

Figure 6: (a): TG/DTG curve of sample SP1. (b): MS spectrum of sample SP1.

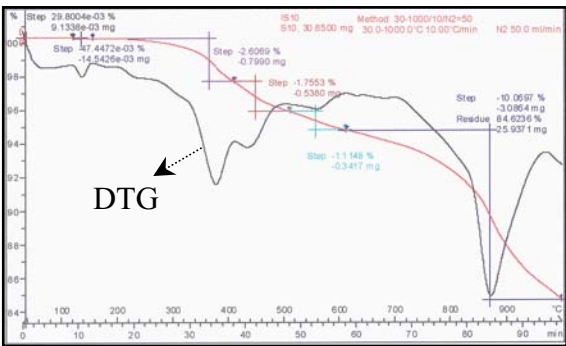


Table 1: MS results showing the possible molecule release.

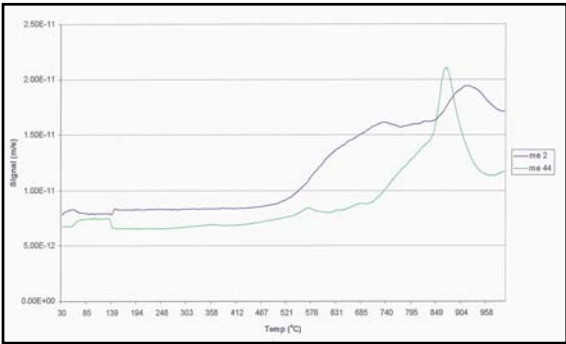
Mass number	Key Fragment	Probable molecule
2	H2+	H ₂
	He++	He
44	C3H8+	C ₃ H ₈
	CO2+	CO ₂
	C2H4OH+	C ₂ H ₅ OH

Mass number 44 also suggests the possible release of propane and ethanol as may be caused by volatilization of organic compounds due to steam at higher temperature. Oxidation of the carbon content might also be possible. This result substantiates the corresponding weight loss between 403°C – 688°C.

Thermal events of sample SP2 is shown in figs. 7(a) and 7(b). Major weight loss occurred in two significant steps beginning from 402°C as in sample SP1 but the total weight loss is slightly higher i.e. 5.48%. MS signal were also detected at these steps with probable molecule as shown in table 1 corresponding to the regions of oxygen inward or the carbon containing species outward.



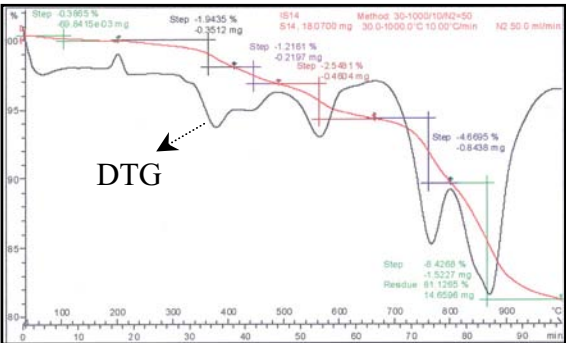
(a)



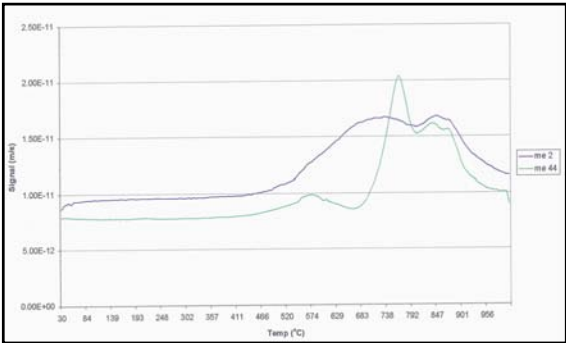
(b)

Figure 7: (a): TG/DTG curve of sample SP2. (b): MS spectrum of sample SP2.

Figs. 8(a) and 8(b) show the TG/DTG curves and MS spectra for sample SP3, respectively. This sample exhibited a 4-step thermal decomposition with a relatively considerable weight loss of 1.94% in the interval of 197°C – 409°C. The corresponding $T_{DTG, min}$ is 374°C, which indicates possible decomposition of the barium element. Two more weight loss steps took place at the DTG temperatures of 439°C and 559°C, with the later emitting MS signals as in table 1. Total weight loss up to 559°C is 5.71%. It is worth mentioning that a significant weight loss of 4.67% was recorded at the 654°C – 789°C temperature zones (probable decomposition of iron oxide) though this is beyond the temperature range of interest.



(a)



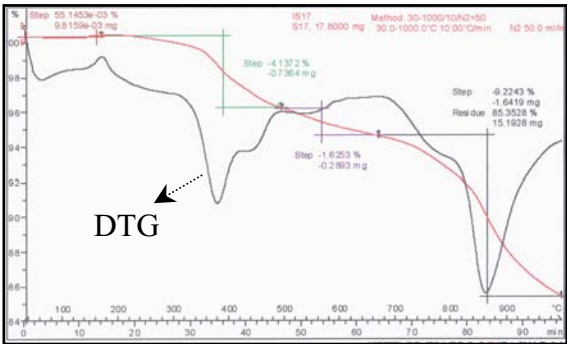
(b)

Figure 8: (a): TG/DTG curve of sample SP3. (b): MS spectrum of sample SP3.

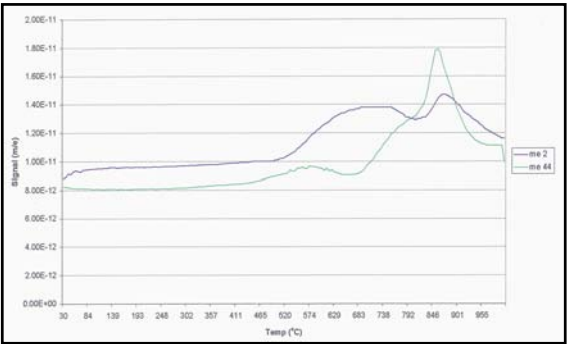
Decomposition of sample SP4 occurred in just three steps as shown in fig. 9(a). Maximum weight loss of 4.14% ($T_{DTG, min} = 378^\circ\text{C}$) was registered at a rather broad temperature range of 171°C – 488°C. The mass to charge ratio of

ionised fragments as shown in fig. 9(b) reveals that gaseous products with molecular mass as in table 1 were released at around 554°C. A weight loss of 1.63% was detected at this temperature zone. Referring to the preceding section, this result indicates the oxidation and/or evaporation of Nipol and barium. The total weight loss at the temperature zone of interest is 5.77%.

Figs. 10(a) and 10(b) are the TG/DTG curves and MS spectra for sample SP5, respectively. The thermogravimetric results show that sample S20 decomposes in four steps up to 653°C. The first two steps produced small reduction in weight totalling 1.2%. Substantial weight loss of 3.58% occurred at $T_{DTG, min} = 378^{\circ}\text{C}$ with discharge of gases as in table 1. This observation is similar to that of sample SP4 but the weight decrease of sample SP5 is lesser by 0.56%, hence an improvement in the formulation. Sample SP5 also registered the lowest total weight loss of 5.30% among all the five samples.

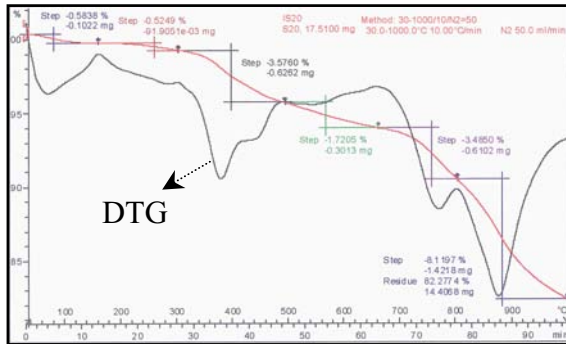


(a)

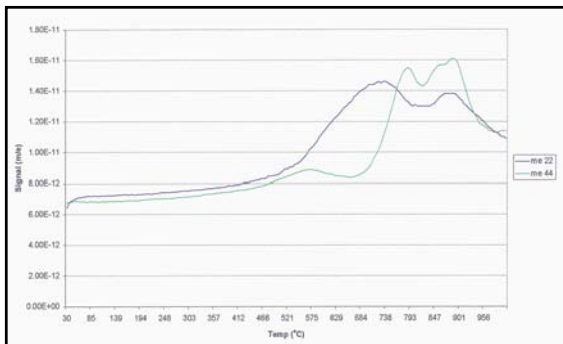


(b)

Figure 9: (a): TG/DTG curve of sample SP4. (b): MS spectrum of sample SP4.



(a)



(b)

Figure 10: (a): TG/DTG curve of sample SP5. (b): MS spectrum of sample SP5.

4 Summary

TG/DTG-MS combined analysis was conducted on five new brake pad formulations. MS results indicated evolution of gases during the temperature ramp especially after 400°C and were related to the physical-chemical reaction of the raw ingredients. Based on the results of TG/DTG, it was found that sample SP1 and SP5 are the favourable samples with the reference to the commercial brake pad. Sample SP1 exhibits a controlled rate of decomposition apparently over a wider temperature range (230°C – 688°C) with a total loss of weight of 5.37%. In contrast, sample SP5 decomposed in a greater magnitude in a single step and narrower temperature gap (303°C – 483°C) albeit the total weight reduction is 5.30%. However, both of these measurements are still high compared to that of commercial brake pad (4.79%) and further refinement to the compositional aspects of the formulations will be undertaken to achieve a stable thermal decomposition and reduced weight loss comparable to the commercially available brake pad.



Acknowledgements

This work is supported by MOSTI Malaysia through IRPA Grant 03-02-01-0055-PR0066/04-03. The authors would like to thank Dr. Mohamad Soib bin Selamat (AMREC, SIRIM), Dr. Talib Ria Jaafar (AMREC, SIRIM) and Dr. Mustafar Sudin (UTP, Tronoh) for their valuable support and useful discussion. Dr. Vengadeswaran and staff members of the Combinatorial Technologies and Catalysts Research Centre (COMBICAT) of University Malaya are thanked for their technical assistance and useful discussions. The administrative support provided by the Faculty of Mechanical Engineering, UiTM and the Institute of Research, Development and Commercialisation, UiTM is appreciated.

References

- [1] Lu, Y., Tang, C.F., & Wright, M.A., Optimization of a commercial brake pad formulation. *J. Applied Polymer Science*, **84** (13), pp. 2498 – 2504, 2002.
- [2] Samyn, P., Schoukens, G., Quintelier, J. & De Baets, P., Friction, wear and material transfer of sintered polyimides sliding against various steel and diamond-like carbon coated surfaces. *Tribology International*, **39**, pp. 575 – 589, 2006.
- [3] Ingo, G.M., Uffizi, M.D., Falso, G., Bultrini & Padeletti, G., Thermal and microchemical investigation of automotive brake pad wear residues. *Thermochimica Acta*, **418**, pp. 61 – 68, 2004.
- [4] Jang, H., Lee, J.S. & Fash, J.W., Compositional effects of the brake friction material on creep groan phenomena. *Wear*, **251**, pp. 1477–1483, 2001.
- [5] Mackin, T.J. et al., Thermal cracking in disc brakes. *Engineering Failure Analysis*, **9**, pp. 63 – 76, 2002.
- [6] Cueva, G., Sinatora, A., Guesser, W.L. & Tschiptschin, A.P., Wear resistance of cast irons used in brake disc rotors. *Wear*, **255**, pp. 1256 – 1260, 2003.
- [7] Mutlu, I., Eldogan, O. & Findik, F., Tribological properties of some phenolic composites suggested for automotive brakes. *Tribology International*, **39**, pp. 317 – 325, 2006.
- [8] Kamruddin, M., Ajikumar, P.K., Dash, S., Tyagi, A.K. & Raj, B., Thermogravimetry-evolved gas analysis-mass spectroscopy system for materials research. *Bulletin of Materials Science*, **26** (4), pp. 449 – 460, 2003.
- [9] Basalik, T., TGA with evolved gas analysis. *American Laboratory*, pp. 24 – 27, 2005.
- [10] Zhang, X., Li, Y., Lv, G., Zuo, Y. & Mu, Y., Thermal and crystallization studies of nano-hydroxyapatite reinforced polyamide 66 biocomposites. *Polymer Degradation and Stability*, **91**, pp. 1202 – 1207, 2006.
- [11] Chiang, I.W. et al., Purification and Characterization of Single-Wall Carbon Nanotubes (SWNTs) Obtained from the Gas-Phase Decomposition of CO (HiPco Process). *J. Phys. Chem. B.*, **105**, pp. 8297 – 8301, 2001.



Section 3

Passenger interface systems

This page intentionally left blank

A fast method for estimating railway passenger flow

Y. Nagasaki, M. Asuka & K. Komaya

Advanced Technology R&D Center, Mitsubishi Electric, Japan

Abstract

To improve the situation for crowded commuters in Japan, it is important to plan a train schedule that considers passenger behavior, such as their choice of trains and the transfer stations used to reach their destinations. However, it is difficult to directly measure such detailed behavior using the present infrastructures, with which we can only get OD (Origin-Destination) data from the automatic ticket gates. The obtained OD data only consists of the number of passengers for each origin-destination and the time each passenger passes through the gates. In this article, to contribute to the planning phase of a new train schedule, the authors propose a method for estimating railway passenger flow using OD data. This paper firstly points out that the problems of estimating passenger flow can be boiled down to a shortest path problem of graph theory by assuming a certain passenger behavior model. By representing train operations in a graph structure, we can assume that a passenger will use the minimum cost path to his/her destination. This paper secondly proposes a method for conducting fast searches of the graph structure. The method uses the fact that railways operate on a time schedule. This method can estimate passenger flow fast enough so as to apply it to a practical train schedule planning support system. Lastly, the authors show the results of applying the passenger flow estimation system to a railway in an urban area in Japan.

Keywords: train schedule evaluation, passenger flow estimation, graph theory, shortest path problem, incremental assignment procedure.

1 Introduction

Currently, planning train schedules depends mainly on expert planners' experiences and intuitions, although a computer based decision support system is gradually appearing. To promote the use of a computer assistance system in a



train schedule planning phase, it is important to evaluate train schedules quantitatively. It is easy for a computerized system to evaluate train schedules from the viewpoint of a railway operator, for example the total travel distance of rolling stocks. However, when experts are planning passenger train schedules, they must consider not only train movements but also passenger behaviors. They must also leave enough transfer time, consider each train's congestion rate, and so on. However, these considerations are very qualitative and do not necessarily suit computerized systems. In order to construct a better decision support system, train schedules must be quantitatively evaluated from the viewpoint of passengers.

To complete such an evaluation, we need detailed information about passenger behaviors such as train choices and the transfer stations used to reach a particular destination. Passenger behaviors depend on a train schedule. We can not obtain such information about passenger behaviors by measuring until trains are operated on the schedule actually. Thus, to evaluate a train schedule in the planning phase, a passenger flow estimation method is required.

In the field of passenger flow estimations, the following methods are currently used.

- A method employing a graph theory's shortest path algorithm [1].
- A method employing network user equilibrium model [2].
- A method employing a logit model [3].

In this paper we propose a fast method for estimating railway passenger flow from a train schedule and OD (Origin-Destination) data. OD data consists of the number of passengers for each origin-destination and the time each passenger passes through the automatic ticket gates. This method is based on a graph theory's shortest path algorithm and it has been improved to work faster than conventional algorithms by using the fact that a railway operates on a time table.

2 Passenger flow estimation using a shortest path algorithm

2.1 The passenger behavior model

From the departing station to the destination station, a passenger can often select his/her route from multiple possibilities. A passenger generally prefers a route that takes the least time, with the least transfer barriers. In this research, we assume that a passenger will select the route that has the least evaluation value. The evaluation value for a route is the sum of the required time of the route and the transfer barrier conversion values. The transfer barrier conversion values represent passenger inconvenience in the dimension of time. That is, we assume that a passenger weighs the inconvenience of having to transfer against the amount of time saved by making a transfer.

2.2 A graph structure representing passenger routes

A graph structure representing passenger routes is created based on a train schedule. The vertex in the graph structure represents each train's departure or arrival at a station. The directed edge from the departure vertex to the arrival



vertex represents the movement of a train between stations. The directed edge from the arrival vertex to the departure vertex represents a train stopping at a station or passengers transferring between trains at a station. Each edge has a cost value that quantifies the degree of passenger inconvenience when choosing the edge. The cost value for a movement edge or a stop edge is the time required of the act represented by the edge. The cost value for the transfer edge is the sum of the time required for the transfer and the conversion value of the transfer barrier. An example of such a graph structure is shown in Figure 1. In such a graph structure, based on the passenger behavior model, a passenger will use the route with the minimum cost from the departure vertex of his/her departing station to the arrival vertex of his/her destination station. Therefore, the shortest path algorithms of the graph theory can be applied to find each passenger's route.

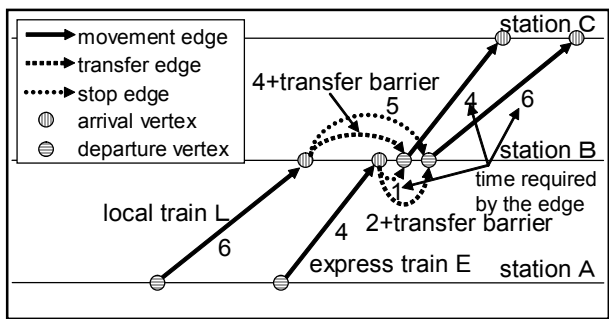


Figure 1: A graph representation of passenger routes.

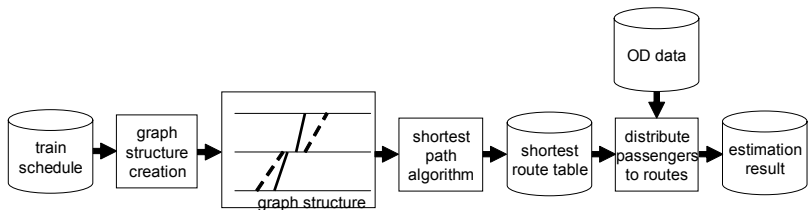


Figure 2: Architecture of passenger flow estimation.

2.3 Applying the shortest path algorithm and estimating passenger flow

Dijkstra's algorithm is often used to find the shortest path from a designated vertex to the other vertices in a graph structure. Previous works simply applied Dijkstra's algorithm on the graph structure that represented passenger routes. We can obtain passenger route tables for any combination of departure and arrival



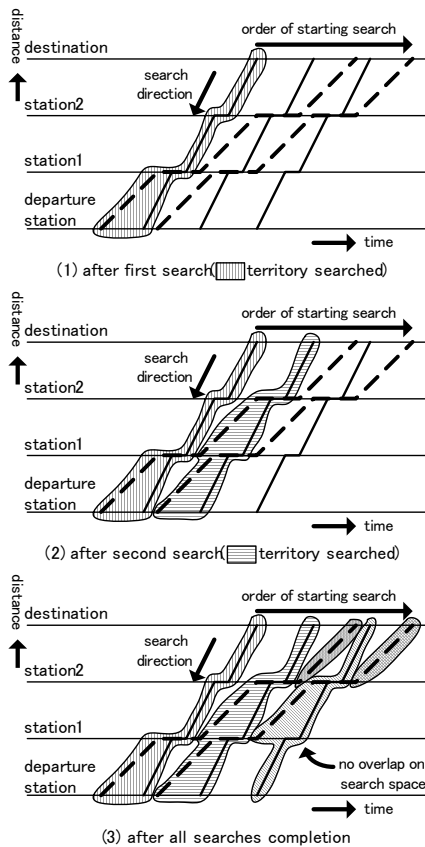


Figure 4: The concept behind the backward preparative search algorithm.

3.3 The backward preparative search algorithm

In the new method, after the graph structure is created, the backward preparative search algorithm is applied to the graph structure. Figure 4 illustrates the concept behind the backward preparative search algorithm. A preparative search starts from the destination station's arrival vertex and tracks back to each directed edge in reverse. From the vertices that are currently searched, a passenger can reach the arrival vertex where the search starts, because a search always tracks back to the directed edges in reverse. The searched vertices are marked to classify the destination station's vertex at the point where the search starts.

Preparative searches are processed in the sequence of destination station's arrival times. A later search should not overwrite nor track back to vertices that are already marked by previous searches. From vertices that are already marked

by previous searches, a passenger can reach an earlier arrival vertex for the destination station, because searches are processed in the sequence of the destination station's arrival time. Therefore, after all preparative searches are completed, each vertex is classified by the earliest arrival vertex of the destination station that a passenger can reach from the vertex.

In a case where the cost of a transfer edge contains the transfer barrier's conversion value, the route by which a passenger can arrive at their destination in the shortest time is sometimes not the most convenient. Thus, the marking on the vertices should be recorded by an evaluation value that shows the best route to a destination. Figure 5 shows vertices with their evaluation values of the best route to a destination. At the beginning of a search, the arrival vertex from the destination station should be marked as 0. When a search tracks back to directed edges in reverse, the evaluation value of the precedent vertex is recalculated by adding the cost of the edge to the evaluation value of the subsequent vertex. If the current evaluation value of the vertex is greater than the value newly calculated, a search can rewrite the evaluation value of the vertex and continuously track back to the precedent edges of the vertex.

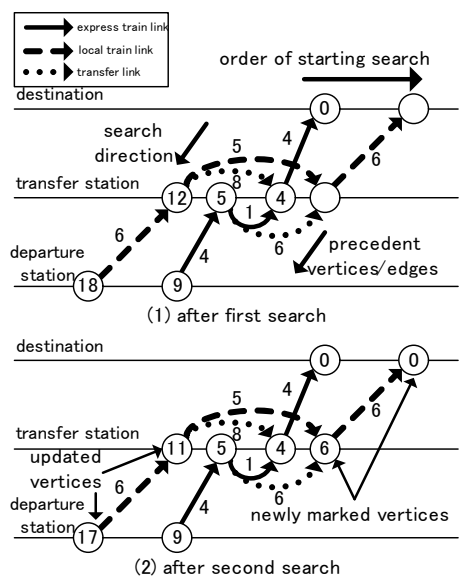


Figure 5: Calculating the evaluation values for vertices using a backward preparative search algorithm.

3.4 The optimized shortest path algorithm

After all backward preparative searches are complete, the optimized shortest path search algorithm is carried out. A search starts from the departure vertex of a departure station and follows the directed edges in a forward direction. During the forward search, the edges that should be searched can be easily identified,



because the evaluation value given to each vertex shows remaining distance to the destination. If the cost value of an edge is not equal to the difference of the evaluation value of both end vertices, the edge doesn't need to be searched. Because the algorithm doesn't follow nonessential edges, it can be processed faster than the simple Dijkstra's algorithm.

The new algorithm proposed in this paper resembles Bellman's Dynamic Programming in that a backward preparative search labels each vertex and later a forward search finds shortest paths. The difference is that the new algorithm uses the features of railway operations to reduce the search space of the preparative backward searches.

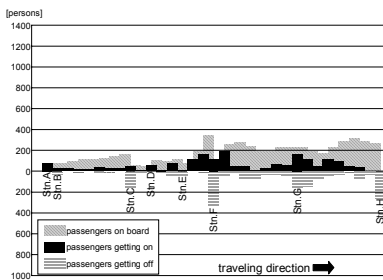


Figure 6: The estimated number of passengers of the local train in the morning.

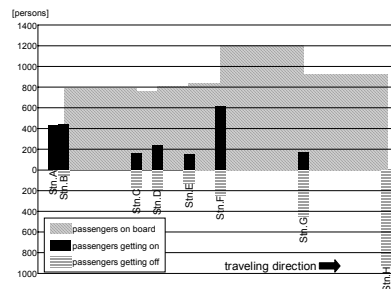


Figure 7: The estimated number of passengers of the express train in the morning.

4 The results of the passenger flow estimation method applied to a railway system in an urban area in Japan

4.1 The estimation parameters

We applied the passenger flow estimation method proposed in this paper to a railway system in an urban area in Japan. The target railway network operates about 1000 passenger trains per day. We used a weekday train schedule and OD data that was taken from automatic ticket gates.

4.2 The estimation results

We executed the estimation on a Pentium 4 3GHz PC; it took about 14 seconds to complete. The conventional method using the simple Dijkstra's algorithm took about 10 minutes against the same data.

Figure 6 and Figure 7 show the estimated number of passengers on the local and express trains in the morning. Stations are shown horizontally and the trains travel from left to right. The upward bar graph from center shows the number of passengers getting on the train at each station. The downward bar graph from



center shows the number of passengers getting off the train at each station. The background bar graph shows the total number of passengers in the train at the time of departure from the station.

5 Congestion-sensitive estimation using the incremental assignment procedure

5.1 The congestion-sensitive estimation method

The passenger behavior model shown in 2.1 assumes that passengers will select their routes according to the journey time and transfer barriers. However some passengers avoid riding on over-crowded trains. Such passenger behavior can be represented by adding an evaluation value for the train congestion to the edge costs. Even so, each train's congestion rate can be calculated only after the passenger flow estimation is complete. To avoid loop references, we use the incremental assignment procedure. The whole passenger demand is divided into several parts. For the first part, the passenger flow estimation is calculated normally. Based on the first result, the costs of edges are changed and passenger flow estimation is calculated for the second part. One by one, the edge costs are changed and the passenger flow estimation is repeated. From this, the total of all these parts gives us the result.

Eqn (1) is a congestion rate to time conversion function taken from [4].

$$f_c'(r) = \max(0.807 \cdot (r/100)^{1.112} - 1, 0). \quad (1)$$

In eqn (1), r means the congestion rate of a train. When the congestion rate is specified, the function returns a coefficient value. If a returned coefficient value is 2, this means that a passenger on a train for one minute will feel as the same as a passenger on an empty train for two minutes.

Figure 8 shows the estimated number of passengers with and without using the incremental assignment procedure in some trains in a section. The number of divisions in the incremental assignment procedure is eight. The illustrated part is during the morning rush-hour, therefore some express trains are very crowded. Using the incremental assignment procedure, some passengers are estimated to avoid riding crowded trains such as Express 2, and use other trains, such as Local trains, instead.

6 Conclusions

In this paper, we have shown a fast method for estimating railway passenger flow. This method can complete estimations faster than the conventional methods using the features of railway operations. This method can also estimate congestion-sensitive passenger flow using the incremental assignment procedure. We can evaluate a train schedule from the viewpoint of passengers by using this method. We can also make train schedules with better services for passengers by calculating the estimated passengers' flow.



References

- [1] Nagasaki, Y., Takano, M., & Koseki, T., Train Rescheduling Evaluation and Assistance System with Passengers' Behavior Simulation Based on Graph Theory, IEEJ Technical Meeting on Systems and Control, SC04-11, pp.25-30, 2004.
- [2] Ieda, H., Akamatsu, T., Takagi, J., & Hatakenaka, H., Commuter's Benefit Evaluation of Train Scheduling by Network User Equilibrium Model, Infrastructure Planning Review No.6, pp.177-184, 1988.
- [3] Hirai, C. & Tomii, N., A Method to Estimate the Number of Passengers for Evaluation of Traffic Rescheduling Plans, RTRI Report Vol.14 No.7, pp.43-48, 2000.
- [4] Yamaga, H. & Hatta, T., Measurement of Commuter's Prostration Cost and Optimal Congestion Charge, The Japan Center for Economic Research Bulletin No.41, pp.110-131, 2000



This page intentionally left blank

Route-choice support system for passengers in the face of unexpected disturbance of train operations

R. Tsuchiya¹, Y. Sugiyama¹, K. Yamauchi¹, K. Fujinami¹,
R. Arisawa² & T. Nakagawa²

¹*Railway Technical Research Institute, Japan*

²*East Japan Railway Company, Japan*

Abstract

When train operation is disturbed due to an accident or natural disaster, passengers have to decide what they should do to get to their destination as early as possible. Unfortunately, however, this is not an easy task for passengers since sufficient information for decision-making is rarely provided to them. Therefore, we have developed a passenger support system which informs passengers of information on the optimal routes to their destinations by taking into consideration the predicted resumption time from the disturbance and the expected required time to each destination. More concretely, the system helps passengers decide whether to take the detour routes to their destinations or wait for the resumption of disturbed operation and continue their journey on the originally scheduled route. The system consists of an optimal route calculation engine and diversified man-machine interfaces, each of which is designed for a specific guidance application. Resumption time and required time to each destination are predicted based upon “resumption process model” whose parameters are determined by the statistical analysis of past data of train operation disturbances. The evaluation test of the prototype system is now being carried out by having a number of subjects in order to prove its effectiveness.

Keywords: route-choice support system, disturbance, passenger guidance.

1 Introduction

In order to increase customer satisfaction, it is very important for railway companies to provide appropriate guidance information to passengers when train



operation is disturbed due to some accidents or natural disasters. When train operation is held up or disturbed, passengers have to decide what they should do to get to their destination as early as possible. Unfortunately, however, this is not an easy task for passengers since sufficient information for decision-making (i.e. time to resumption, estimated time to destination, etc) is rarely provided to them. Therefore, we have developed a route-choice support system which informs passengers of the optimal routes to their destinations by taking into consideration the predicted resumption time from the disturbance and the expected required time to each destination.

In the remainder of this paper, we will make an account of the background of this research, basic concepts of the system, followed by the design and implementation of the system. Also included is the result of the questionnaire survey which has been carried out by having a number of respondents randomly sampled from commuters in the urban area of Tokyo.

2 Background of the research

In Japan, there are several commercially available route-choice support systems for public transportation networks, which provide users with optimal travel plans to go to a specified destination. Few of them, however, can assist passengers with their decision-making when railway operations are disturbed. Some researchers of RTRI (Railway Technical Research Institute) have already proposed a passenger support system which calculates optimal detour routes to the destination when some of the railway lines are suspended. Although this feature is already supported by some of the route search engines commercially available, they can not help passengers decide whether they should take the detour route to the destination, or wait for the resumption of the suspended line and continue their journey on the originally scheduled route. In order to respond to these needs of passengers, more sophisticated decision support system is necessary.

Whether passengers can make appropriate decision depends on the correctness of the prediction as to when train operations will be resumed as well as the knowledge on the train operation rescheduling process. Therefore, we have developed a route-choice support system which informs passengers of information on the optimal routes to their destinations by taking into consideration the predicted resumption time from the disturbance and the estimated time to destination.

3 Route-choice support system

3.1 Estimation of required time to destination

In this research, we assumed that passengers in the face of unexpected disturbance of train operations want to know the fastest route to destination (figure 1). Based on this assumption, we propose a method which effectively realizes information provision that meets the demands of passengers.



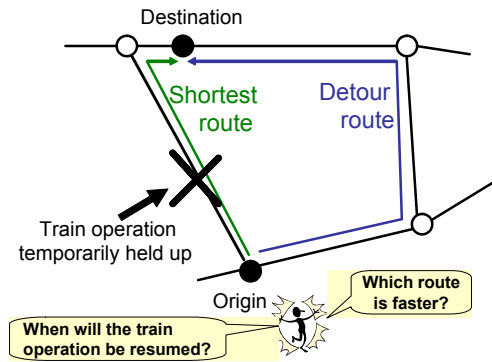


Figure 1: Passengers want to know which route is faster when train operations are disturbed.

We have devised a simplified resumption process model to estimate the required time between two adjacent stations. Required time is approximated by linear function as is depicted in figure 2.

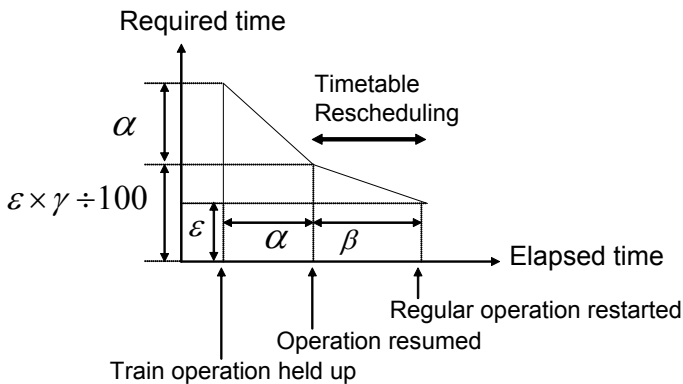


Figure 2: Simplified resumption process model for estimating required time between two adjacent stations.

α represents the elapsed time from occurrence of disturbance to the resumption of operation, whereas β represents the elapsed time from the resumption of operation to the regular operation (i.e. operation with regular interval). Please note that even when the train operation is resumed, it has to undergo train rescheduling process, during which it often takes more time for passengers to go to their destinations than when trains are in regular operation. This is mainly due to the fact that in train operation rescheduling process, average speed and intervals of trains are often reduced. γ represents the ratio

which indicates how much extra time is needed in order to go from one station to another when train operation is in the rescheduling process. In our resumption process model, the value of each parameter is determined base on the statistical analysis of past data of train operation disturbances. These values are default values which can be modified by human operators.

3.2 Presenting guidance information to many and unspecified passengers

Guidance information can be provided to passengers in two different ways. One method is to present guidance information to many and unspecified passengers, typically at stations by using large screen displays. Another method is to deliver personalized guidance messages to passengers’ portable terminals including cellular phones. One of the difficulties in presenting guidance information to many and unspecified passengers is that unlike personalized information provision, it is impossible to customize the presentation based on a passenger’s destination, his/her knowledge level on the detour route, his/her ability to use electronic devices, etc. Usability of the system heavily depends on the way how the guidance information is presented to passengers and how effective it is when they make travel decisions. In this research, therefore, we have devised two different user interfaces for presenting guidance information for many and unspecified passengers, which will be evaluated to find the best way. One user interface which we have developed is based on the map of railway network for urban area of Tokyo, which we often see at stations or on board trains. Appropriateness for taking a detour route, which depends on the destination, is classified by colour per station on the map (figure 3).

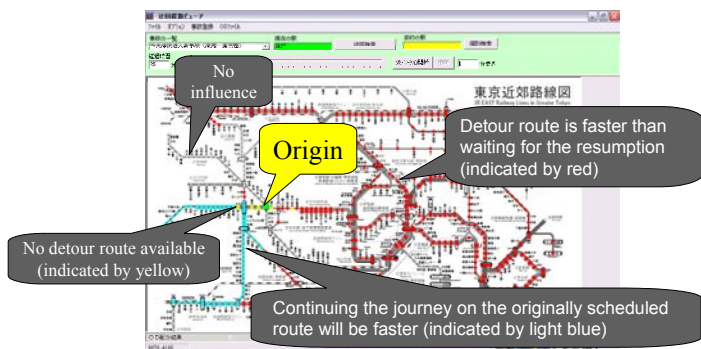


Figure 3: Appropriateness for taking a detour route.

As is shown in figure 3, every station is attached one of the four colours (i.e. red, yellow, light blue and colourless). The configuration of colours changes every minute so that up-to-date information is presented to passengers.

Table 1 briefly describes the meaning of colours attached to each station.

Table 1: Colors and their meanings.

categories	colors	meanings
Detour route recommended	red	Detour route is advantageous over originally scheduled route
No detour route	yellow	No detour route available. You have no choice but wait for the resumption
Originally scheduled route will be better	light blue	Detour route available but waiting for resumption and continue the journey on the originally scheduled route will be better
No influence	colorless	No influence of disturbance

“Detour route recommended” (red) indicates that the destination (station) can be reached earlier by taking detour route than by originally scheduled route. “No detour route” (yellow) indicates that there are no detour routes to the destination and passengers have to wait for the resumption of the operation and continue their journey on the originally scheduled route. “Wait for the resumption” (light blue) indicates that passengers will be able to arrive at the destination earlier by waiting for the resumption and take the originally scheduled route than by taking the detour route.

This type of presentation can inform passengers whether taking the detour route is advantageous over taking the originally scheduled route by waiting for resumption. It does not, however, tell passengers how they can go to the destination by using the detour route. Therefore we have devised a way of presenting more detailed information on the detour route for restricted number of destinations so that passengers going to those major destinations will be provided with more detailed guidance (figure 4).

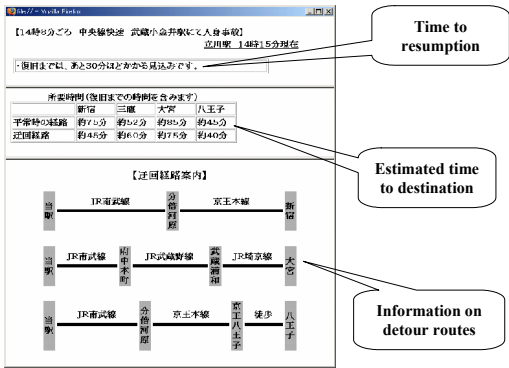


Figure 4: Screen which indicates estimated time to destination and detour route details.

3.3 Application to personalized guidance

Although our main target is information provision for many and unspecified passengers, our method can also be applied to personalized guidance. Figure 5 is an example of messages delivered to a portable terminal which includes time to resumption, estimated time to destination as well as detour route details [1].

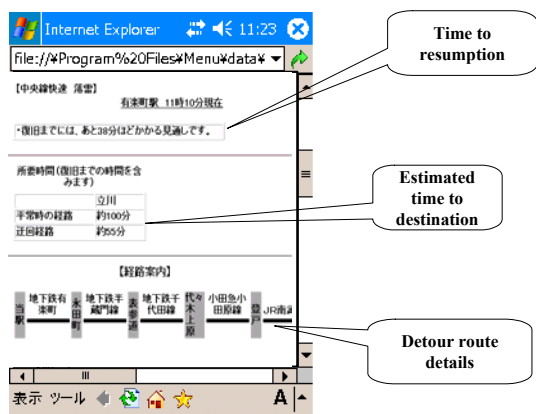


Figure 5: Example of guidance message delivered to portable terminal.

3.4 System implementation

The system that we have developed consists of optimal route calculation engine and several applications which, by cooperating with the engine, present guidance information in a variety of ways. Typical application of the system might include, as is described in previous sections, variable signs for many and unspecified passengers at stations, support systems for station staffs engaged in passenger guidance as well as personalized guidance systems using passengers' portable terminals. Therefore we have developed a general purpose optimal route calculation engine which include detour route calculation feature (figure 6).

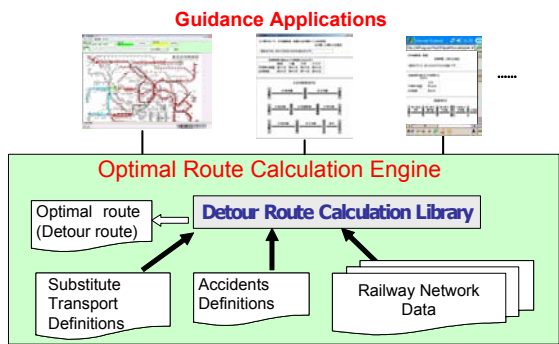


Figure 6: Optimal route calculation engine.

Core functionalities of the system are realized by the detour route calculation library which calculates optimal routes from a specified station to all JR (Japan Rail) stations, taking into consideration disturbances of railway operation. Input data to the system consists of railway network data, substitutable transport definitions and accidents definitions. Railway network data covers nation-wide JR lines as well as lines belonging to all private railways. In Japan, most of the railway companies are in agreement with each other on the mutual provision of substitute transport in case some of the railway companies encounter disturbance of train operations and therefore cannot provide sufficient transport capacity. Since these requests are usually accepted, passengers can take detour routes without any extra charge. Input data “substitution transport definitions” represents basic patterns which specify which railway lines should be used as substitution transport depending on the place where a train operation disturbance takes place.

Input data “accidents definitions” defines parameters (i.e. α, β, γ) related to accidents which take place in some of the JR lines (figure 7). Although the default values of parameters are determined by statistical analysis of past data of train operations disturbances, users (i.e. commanders or station staffs of each railway) can modify the values of parameters when necessary.

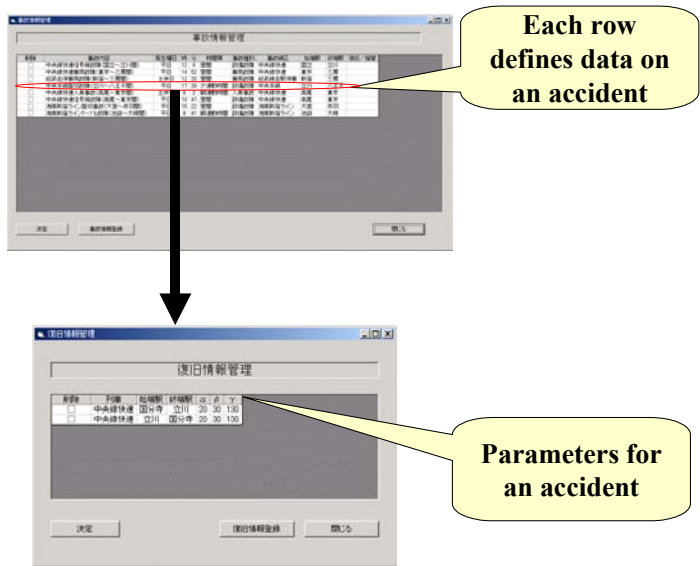


Figure 7: Accidents data definitions.

In order to find the optimal (i.e. shortest) route to a destination, the system has to calculate expected time to destination for several candidate routes. In our system, the expected time to a destination via a specified route is calculated by accumulating the expected time between two adjacent stations which belong to the route.



If the candidate route does not contain disturbed railway lines, the weight of the link which correspond to a pair of adjacent stations are specified based on the average running time between these two stations. Whereas, when the candidate route contains disturbed railway lines, the system estimate the running time between the stations based on the resumption process model depicted in figure 2 and set the corresponding value as the weight of the link. When calculating the estimated running time between two adjacent stations, the system takes into consideration the elapsed time to get to either of the two stations which is also estimated by the model. For example, if you are going from a station which is far away enough from the disturbed railway line, there will be little effect on your itinerary even if it contains the disturbed line.

4 Evaluation

The optimal-route calculation engine which we have developed can calculate optimal routes from a specified station to all JR stations in Japan in a few seconds by using a general purpose office PC. The performance that we have achieved with our calculation engine is sufficiently high to realize the information update interval which we expect to be one minute.

We have so far demonstrated our system at several expositions and rail-related conferences and received favorable reactions from participants. They highly evaluated its effectiveness as a tool to help passengers' decision-making when they encounter train operation disturbances. With the increasing needs of information provision for passengers when train operations are disturbed, several railway companies are currently trying to provide passengers with more information in a timely manner even if it contains uncertainty. For example, some of the railway companies in Tokyo urban areas started to provide information related to the time when the train operation is expected to be resumed. Although this kind of information contains some uncertainty and is not always true, they seem to be accepted favorably by most of the passengers.

Providing information timely even though it contains uncertainty may be justified by the questionnaire survey which we have carried out as a part of this research. We asked the respondents about their attitudes toward the provision of uncertain information as to when the train operations are expected to be resumed. More concretely, we asked how similar their attitudes are to the following two contrastive attitudes.

A: I expect that information be provided timely, even though it contains some uncertainty

B: I expect that information be provided only when it is certain. Uncertain information may cause confusion.

According to the result of the survey, 94% of the respondents' attitudes are similar to that of A.

We are currently carrying out the test of the prototype system by having a number of subjects randomly sampled from commuters in Tokyo urban area in order to evaluate the effectiveness of the system, appropriateness of presentation as well as its acceptability as a tool to help passengers' decision-making.



5 Conclusions

We have developed a route-choice support system which helps passengers make appropriate decision when train operations are disturbed. The system helps passengers decide whether to take the detour routes to their destinations or wait for the resumption of disturbed operation and continue their journey on the originally scheduled route. The system consists of an optimal route calculation engine and diversified man-machine interfaces, each of which is designed for a specific guidance application. Typical application of the system might include variable signs for many and unspecified passengers at stations, support systems for station staffs engaged in passenger guidance, personalized guidance systems using passengers' portable terminals, etc.

We have so far demonstrated our system at several expositions and rail-related conferences and received favorable reactions. They highly evaluated its effectiveness as a tool to help passengers' decision-making when they encounter train operation disturbances.

The evaluation test of the prototype system is now being carried out by having a number of subjects and its result will be reported soon.

References

- [1] Tsuchiya, R., Nozue, M.: "Supporting Intermodal Travelers by Agent-based Information Integration," 8th World Congress on Railway Research (WCRR 2006), 2006 (to appear)
- [2] Tsuchiya, R., Sugiyama, Y., Yamauchi, K., Fujinami, K., Arisawa, R., Nakagawa, T.: "Route-choice Support System for Passengers Based on Predicted Resumption Time from Unexpected Disturbance of Train Operations," RTRI Report, Vol. 20, No.2, Feb. 2006 (in Japanese)



This page intentionally left blank

A new delay forecasting system for the Passenger Information Control system (PIC) of the Tokaido-Sanyo Shinkansen

K. Fukami, H. Yamamoto, T. Hatanaka & T. Terada
Central Japan Railway Company, Japan

Abstract

The former delay forecasting system for the PIC calculates the delay by adding a fixed station-to-station running time to the departure time. However the system has two problems. The first problem is that it cannot reflect the delay that occurs between stations, which is very important because the distance between stations of the Tokaido-Sanyo Shinkansen is very long. The second problem is that the delay often changes until the train departs the terminal station owing to transportation changes by the conductors. To solve these problems, we have developed a new delay forecasting system for the PIC of the Tokaido-Sanyo Shinkansen that simulates actual train movement.

The system mentioned above has been used since July 2003. This system has offered precise delay information since then, and has made great contributions towards improving the service level of Tokaido-Sanyo Shinkansen.

Keywords: simulation, passenger interface, traffic control.

1 Introduction

The Tokaido-Sanyo Shinkansen, the main artery of the Japanese transport network, is a very punctual railway system. When delays occur, passengers expect precise delay information. To satisfy such demand, PIC reports delay of the trains to departure indicators and automatic announcement systems at stations. To improve the service level of Tokaido-Sanyo Shinkansen, we have developed a new delay forecasting system for the PIC that simulates actual train movement.



2 Features of the new delay forecasting system for the PIC

2.1 Simulation of actual train movement

The former delay forecasting system for the PIC calculates the delay by adding fixed station-to-station running time to the departure time. However, it cannot reflect the delay that occurs between stations, which is very important because the distance between stations of the Tokaido-Sanyo Shinkansen is very long.

To solve this problem, we have developed a new delay forecasting system for the PIC that simulates actual train movement and reflects the delay between stations. Using the train tracking information, the system calculates the current position and speed of all trains that exist in Tokaido-Sanyo Shinkansen. On the basis of the information, it simulates actual train movement and generates the forecasted diagram and delay every 15 seconds.

Fig.1 shows an example of the guidance when the train stops between stations. The former delay forecasting system cannot indicate the delay even if the train delays between stations and passengers have to wait for the train on the platform. On the other hand, the new delay forecasting system can indicate the delay between stations, and passengers are able to wait in a comfortable waiting room.

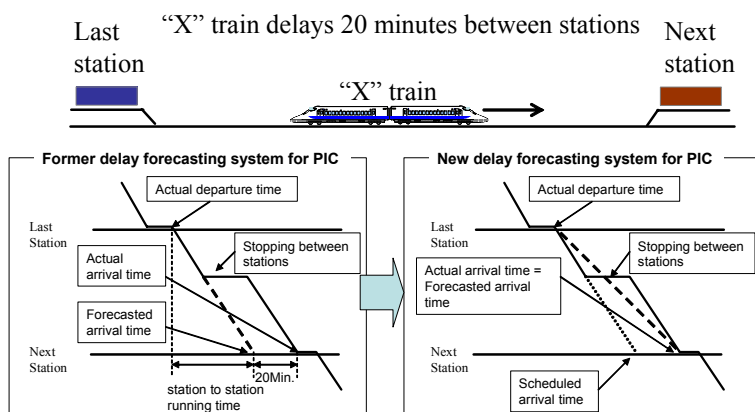


Figure 1: Example of the guidance when the train stops between stations.

2.2 Generation of delay status information

The former delay forecasting system calculates the delay of shuttle train using the previous operating train and transmits only the delay to the PIC. However, when the delay occurs, conductors modify diagram and car scheduling to minimize the delay. It causes a decrease in the delay of the shuttle train, and if passengers believe the delay indicated before the modification, they are likely to miss their reserved train.

To solve the problem, the new system generates the delay status information as well as the delay. Using this information, departure indicators do not indicate



the real delay, but the delay status information such as “delay expected”, until the shuttle train departs the origin station.

Fig.2 shows the example of the shuttle train’s guidance. The situation is as follows. ‘Nozomi 8’ delays and the shuttle train ‘Nozomi 21’ expects to delay 30 minutes at Shin-yokohama station. After conductors modify car scheduling from ‘Nozomi 8’ -> ‘Nozomi 21’ to ‘Deadhead 9890’ -> ‘Nozomi 21’, the delay of ‘Nozomi 21’ decreases to 0 minutes. In this case, the former delay forecasting system indicates that the delay of ‘Nozomi 21’ changes from “about 30 min” to “0 min”. If passengers who reserve ‘Nozomi 21’ believe the indication “about 30 min”, they are likely to miss the train. On the other hand, the new delay forecasting system does not shows the concrete delay until the shuttle train departs the origin station. In this case, it indicates the delay that changes from “delay expected” to “0 min”, and it does not lead the passengers to miss the train.

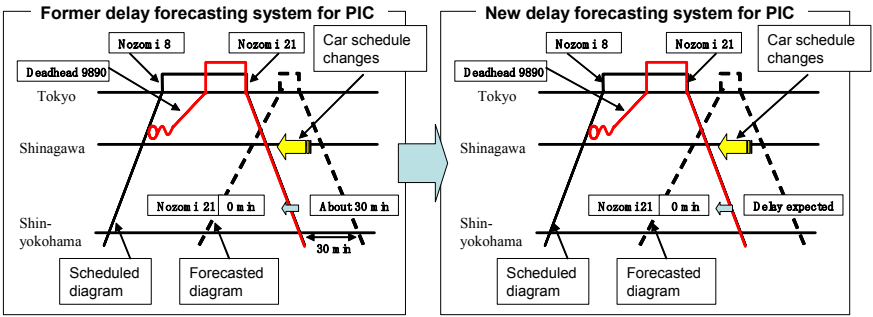


Figure 2: Example of the shuttle train’s guidance.

3 System configuration

3.1 System Architecture

Fig.3 shows the system architecture of the new delay forecasting system for the PIC. Former delay forecasting system exists in EDP (Electric Data Processing computer) system. However, the new delay forecasting system needs train tracking information which is managed in the PRC (Programmed Route Control) system. Therefore, the delay forecasting system is developed in PRC system, and to influence the existing function, we employed a new workstation called DFP (Diagram Forecasting Processor). To enhance the reliability, we adopted dual system to the DFP.

3.2 Software Architecture

The new diagram forecasting system consists of 3 functions, “Train tracking function”, “Current condition detecting function” and “Train movement simulation function”.



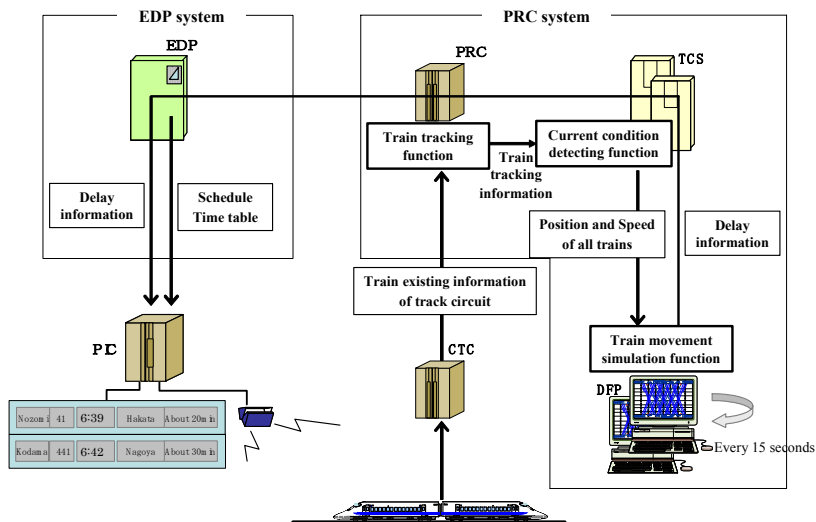


Figure 3: System configuration.

3.2.1 Train tracking function

Train tracking function in PRC tracks the trains using the track circuits' sequential ON/OFF movement, which is obtained from CTC (Centralized Traffic Control System). This function renews the train tracking information and sends it to the current condition detecting function every time the function receives information from CTC.

3.2.2 Current condition detecting function

Current condition detecting function in TCS (Traffic Communication Server) calculates the speed and the position of all trains every time the train tracking information as follows.

First, it calculates ATC (Automatic Train Control system) speed signal by the existence track of previous train. After that, the speed of all trains is calculated:

1) If the train exists in the border of a track circuit, the speed of the train is calculated as follows:

Speed = the length of the train / the time between ON/OFF movement of the track circuit

2) If the train exists in the middle of a track circuit, the speed of the train is calculated as follows:

a) The target speed is calculated, using minimum of ATC speed signals, speed limit and maximum speed of the car.

b) The train's acceleration or deceleration is decided by comparing the previous speed with the target speed.

c) The speed of the train is calculated by the decision of the b) and the previous speed.

3) Finally it renews the position of all trains, using the calculated speed above, and sends it to the train movement simulation function.



3.2.3 Train movement simulation function

Train movement simulation function in DFP simulates the train movement on the basis of the current speed and position, and makes the forecasted diagram and the delay of all existing trains. This function simulates the train movement reflecting all conditions such as the train performance, ATC signal speed, slope of the position, speed limits and so on. To reflect the delay between stations, it is sufficient to forecast using the train movement simulation up to the next station. In order to reduce CPU load, after this function simulates the next station's delay using the train movement simulation, it calculates the next 5 stations' delay using the conventional method, which adds station-to-station running time that reflects temporary speed limits.

This function also generates delay status information on the basis of the position of the trains and car scheduling.

Every 15 seconds, the function generates the delay and delay status information, and they are transmitted to the PIC which controls departure indicators and automatic announcement systems at stations by way of TCS, PRC and EDP.

4 Conclusion

The information of the PIC is very important, and the position and the speed of trains differ every time, therefore we had to test the system for more than a year. In order to improve the quality, we have also developed some debugging tools that replay the conditions, display the running profile, detect the change of the delay, etc.

The system mentioned above has been used since July 2003. Owing to sufficient tests, this system has offered precise delay to the passengers since then. It has made great contributions towards improving the service level of Tokaido-Sanyo Shinkansen.

Reference

- [1] Fukami, K. & Yamamoto, H., A new diagram forecasting system for the Tokaido-Sanyo Shinkansen. *World Congress on Railway Research*, Koln, Germany, No 86, 2001.



This page intentionally left blank

Section 4

Decision support systems

This page intentionally left blank

Reconstruction of train trajectories from track occupation data to determine the effects of a Driver Information System

T. Albrecht^{1,2}, R. M. P. Goverde¹, V. A. Weeda¹ & J. van Luipen³

¹*Faculty of Civil Engineering and Geosciences,
Delft University of Technology, The Netherlands*

²*“F. List” Faculty of Traffic and Transportation Sciences,
Dresden University of Technology, Germany*

³*Department Strategy & Innovation, ProRail, Utrecht, The Netherlands*

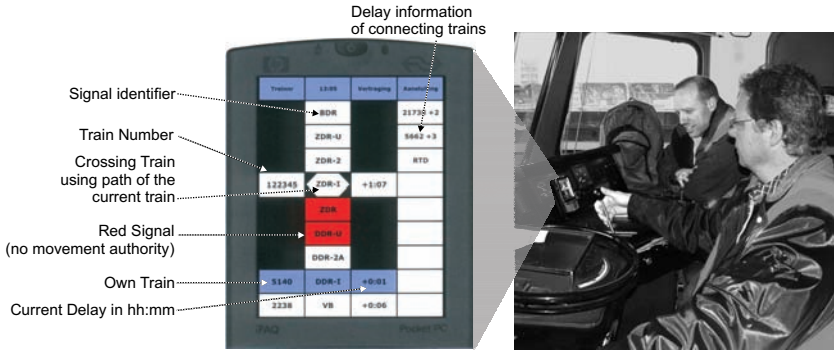
Abstract

The Dutch railway network is one of the most intensely utilized networks in the world and therefore many conflicts occur. The infrastructure manager ProRail developed and tested a Driver Information System (DIS), which gives information on current positions and delays of other trains on the same track. In a potential conflict situation, the driver is then able to anticipate to the behaviour of the previous train which should help avoiding unscheduled stops before red signals and reducing the consequences of a conflict (increased energy consumption and delay).

During the real-world test of the DIS, only few trains were equipped with accessible onboard measurement equipment which was insufficient for a profound analysis of the effects of it. It was therefore decided to reconstruct train trajectories by using track occupation data, which is available for all the trains with a precision of one second. A least squares approach is proposed to obtain the most probable train speed profile along the line. The reconstructed trajectories are used to support the manual observations of special events during the tests.

Keywords: driver information, traffic management, train detection system, speed measurement.





Although timetables are constructed conflict-free with regard to schedule d train paths, it is impossible to avoid conflicts in practice due to unforeseeable random disturbances. Conflicts often lead to unplanned stops which cause delays to the trains that were forced to stop and also consecutive delays to other trains using the same route. Moreover, this also results in a higher energy consumption of the concerned trains due to unplanned re-acceleration.

In order to reduce the number and the consequences of those conflict situations, the Dutch railway infrastructure provider ProRail developed a Driver Information System (DIS) called RouteLint within the research project “Het Spoor Meester” (Dutch for: The Master of the Rail). This DIS displays train positions (per signal section) and delays of other trains as well as the already reserved path of the own train, cf. Fig. 1. With this additional information and their professional skills the drivers should be able to anticipate to conflicts before they occur and to adapt their way of driving in order to decrease the negative effects of conflict situations.

After very promising simulator tests of this system [1], it was decided together with the most important Dutch railway operators (NS Reizigers and NS Internationaal for the passenger trains and Railion for freight trains) to test the system in practical operation on the heavily utilized line between Rotterdam and Dordrecht during two weeks in November/December 2005. Before the tests, some train drivers were specially instructed using the DIS to be ambassadors of the project. During the tests, those ambassadors got onboard of selected trains travelling on the corridor and had to interpret the information given by the DIS to the driver in charge of the train (Fig. 1). Although the number of trains equipped with RouteLint was limited by the number of available ambassadors, more than 100 test runs have been done.

1.2 Measuring the effects

During the time the ambassadors were on board, they noted interesting events connected with the additional information given by the DIS.

It was nevertheless necessary to support this subjective information by objective data, e.g. train speed profiles or energy consumption. However, energy consumption is not measured per train but in the power substations. Due to the relatively low number of trains equipped with DIS it seemed impossible to use this data for an estimation of the effects of the system. The experiment outline (rolling stock equipped with DIS not known in advance) also did not allow to install onboard devices especially for the trials and neither were the data of the black boxes (train run recorders) accessible for non-safety purposes.

To overcome this dilemma Delft University of Technology proposed to make use of available track occupation times to reconstruct train speed profiles as suggested earlier by Goverde [2]. Those trajectories could partially be validated by measurements from energy meters installed onboard of a few trains.

This paper describes the theoretical background of the proposed method in section 2 and its practical application in section 3, which also discusses issues of data procurement, some sample trajectories and the validation of the approach with the energy meter data. Section 4 summarizes main results and gives an outlook on future research.

2 Theoretical background

2.1 Speed profile estimation

The motion of a train is a dynamic process with three variables that change over time: the position of a train, its speed, and the acceleration (or deceleration for negative values). In the sequel, it is assumed that the section passage times and section lengths are available. The passage times of sections on a route are generated by the VTL-tool, a tool similar to TNV-Prepare [2, 3]. The section lengths are provided by ProRail.

Consider a train run over a route of n track sections and assume that the n section entrance times are available from track circuit occupancy measurements. The train dynamics on this route can be described by the discrete Newton laws of motion

$$\begin{cases} x_{k+1} &= 3D \quad x_k + v_k(t_{k+1} - t_k) + \frac{1}{2}a_k(t_{k+1} - t_k)^2 \\ v_{k+1} &= 3D \quad v_k + a_k(t_{k+1} - t_k), \end{cases} \quad (1)$$

for $k = 3D1, \dots, n-1$. Here x_k is the (known) route position of the beginning of section k , v_k is the speed at this section boundary, a_k is the (average) acceleration on section k , and t_k is the measured entrance time of section k . The main problem here is that the measured passage times are rounded to seconds and therefore the



passage times are unknowns within some bounds. Problem (1) can be solved using a least-squares approach as explained next.

Let τ_k denote the realized entrance time of section k (in seconds). Hence, measurement t_k is just τ_k rounded down in seconds and so $\tau_k \in [t_k, t_k + 1)$. Then the problem is finding variables v_k ($k = 3D1, \dots, n$), a_k ($k = 3D1, \dots, n - 1$), and τ_k ($k = 3D1, \dots, n$) with $\tau_k \in [t_k, t_k + 1)$ such that

$$J_k(v_k, a_k, \tau_k, \tau_{k+1}; x_k) := 3Dx_k + v_k(\tau_{k+1} - \tau_k) + \frac{1}{2}a_k(\tau_{k+1} - \tau_k)^2$$

is equal to x_{k+1} for all $k = 3D1, \dots, n - 1$. Note that x_k is a known parameter in $J_k(\cdot)$ and not a variable. Posed this way, the problem has some redundancy: it is sufficient to find the initial speed v_1 and actual passage time τ_k at the first section, and then subsequently in each step k the appropriate variables a_k and τ_{k+1} , which then also fixes the speed v_{k+1} by the second equation in (1). The decision variables can therefore be reduced to the $2n$ variables $(\tau_1, \dots, \tau_n, v_1, a_1, \dots, a_{n-1})$.

The problem can now be defined as a least-squares optimization problem that minimizes the sum of squared errors of $x_{k+1} - J_k(v_k, a_k, \tau_k, \tau_{k+1})$, while satisfying $\tau_k \in [t_k, t_k + 1)$ for all k . Instead of explicitly adding the passage time constraints to the optimization problem it is more convenient from an algorithmic point of view to extend the objective function with additional terms that penalize large deviations from the actual passage times. Consider therefore the following least-squares optimization problem

$$\begin{aligned} \text{Minimize} \quad & \sum_{k=3D1}^{n-1} (w_1 (x_{k+1} - J_k(v_k, a_k, \tau_k, \tau_{k+1})))^2 \\ & + \sum_{k=3D1}^n \left(w_2 \left(\tau_k - t_k - \frac{1}{2} \right) \right)^2 + \sum_{k=3D1}^{n-2} (w_3 (a_{k+1} - a_k))^2 \quad (2) \end{aligned}$$

subject to $v_{k+1} = 3Dv_k + a_k(\tau_{k+1} - \tau_k)$ for $k = 3D1, \dots, n - 1$. This constraint implicitly determines the speeds v_k ($k = 3D2, \dots, n$) in terms of passage times and acceleration, and only serves for notational convenience. In the actual implementation these (equality) transformations are substituted directly in the cost function. The first term in the cost function penalizes the deviation of the section position x_{k+1} and the estimated train position at time τ_{k+1} given by $J_k(v_k, a_k, \tau_k, \tau_{k+1})$. The second term penalizes the deviation of the actual passage time τ_k from the expected passage time $t_k + 1/2$. Recall that a rounded measurement t_k is in fact a uniformly distributed stochastic variable on $[t_k, t_k + 1)$ with mean $t_k + 1/2$. The third term models that jumps in acceleration or deceleration are as small as possible over subsequent sections and is added to direct the least-squares algorithm towards satisfying solutions. The coefficients w_1, w_2 and w_3 are weight factors to balance the various terms. Note that an absolute deviation of 1 s of the passage time estimates τ_k from the measurements is allowed. However, a jump of 1 m/s² over subsequent accelerations is unacceptable. Therefore the acceleration jump terms can be multiplied by a weight of for instance



$w_3 = 3D100$. In this case a change in value of a passage time τ_k of 1 second is in the same order as a change of 0.01 m/s^2 in acceleration jump (assuming $w_2 = 3D1$). The error of the first term should be zero and therefore w_1 should be large compared to the other two weights.

2.2 Limitations of the proposed methodology

2.2.1 Measurement precision

Track circuit occupation times are rounded down to the full second. This means that the actual time difference between two adjacent sections, which determines the average speed on this section, may (in the extremes) be a second shorter or longer than the measured time difference. This has special importance on sections with very short occupation times, e.g. very short sections or sections passed with high speed.

With current train detection equipment available on the market, a higher precision can be achieved. It is however questionable whether this improvement can also be used in practice at reasonable cost or whether other sources of errors (e.g. transmission delays) will then dominate the quality of the data.

2.2.2 Assumption of a single acceleration rate per section

In the proposed approach, a single acceleration rate per section is assumed. In particular, it is assumed that a train does not stop at intermediate sections. In some situations this assumption causes severe instability to the filtering algorithm, one of which is illustrated in Fig. 2a. Here, a train slows down with constant deceleration rate at the beginning of section 2 and comes to a complete standstill, which corresponds to a second acceleration rate of 0 within this section.

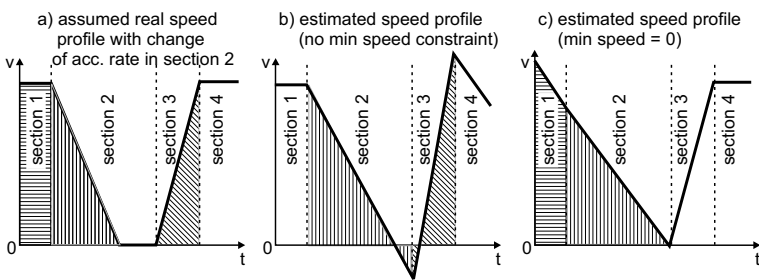


Figure 2: Three speed profiles which all fulfill the constraints of constant section lengths (areas under the curves) and passage times at section borders.

The algorithm then tries to match section occupation times and section lengths to a train behaviour with only one acceleration rate. Part b) and c) of Fig. 2 show solutions, which could have been produced by the algorithm, but do not correspond to the real train behaviour: In order to match section lengths and occupation times with only one acceleration rate, solution b) proposes negative speeds. If one tries



to avoid these irregular values by adding a penalty for negative speeds to the cost function, the estimated speed before the section with two acceleration rates will diverge significantly from the real train behaviour (cf. Fig. 2c). The wrong estimations can moreover influence sections close to the area where the switching occurs, although the oscillations generally become smaller at larger distance.

Choosing a higher weight factor for acceleration changes w_3 , the algorithm can find solutions, which do not contain those strong oscillations. As a consequence, the obtained solution will most probably have higher (and sometimes negative) time differences (term 2 in eq. 2).

The shorter a section is, the less is the probability of a change of regime within this section, so the less the influence of this error becomes.

3 Application

3.1 Test line

The RouteLint DIS was tested on the four-track railway line between Rotterdam and Dordrecht. This 20 km long line is used by six intercity and international lines and three regional train lines and has an important number of freight trains due to the connection to the Kijfhoek freight terminal which is the most important of its kind in the Netherlands. There are two important elevation changes on this line: The Willemspoor tunnel near Rotterdam and the bridge over the Oude Maas just before Dordrecht, which both have gradients up to 17‰.

The whole line comprises more than 400 track circuits (including different variants around switches/ track crossings). In the larger stations, the average length of track circuits is 110 m, while on the open track (including 5 minor stations and several junctions to the freight terminal) this is around 320 m. The shortest track circuit length was 25 m and some aggregated sections were up to 1000 m length.

3.2 Software design

In order to be able to efficiently compute different speed profiles, a dedicated software tool was developed in MATLAB to

1. integrate all the different data sources and prepare the filtering,
2. filter the data according to the algorithm described in sect. 2.1,
3. output major results graphically and as aggregated data in html-format.

The information flows between the different data sources are illustrated in Fig. 3. Data procurement as well as logically connecting the different data sources represented the main part of the work. For instance, track circuit positions had to be identified manually from plans because they were not available elsewhere, including all the different paths at track switches and crossings. Gradient data was available digitally, but had to be matched to the circuit positions via a track number, because different tracks had a different height profile on the examined line. Finally, sections had to be identified, where trains were allowed to stop, because the filtering algorithm can not work correctly if stops are not correctly identified due to



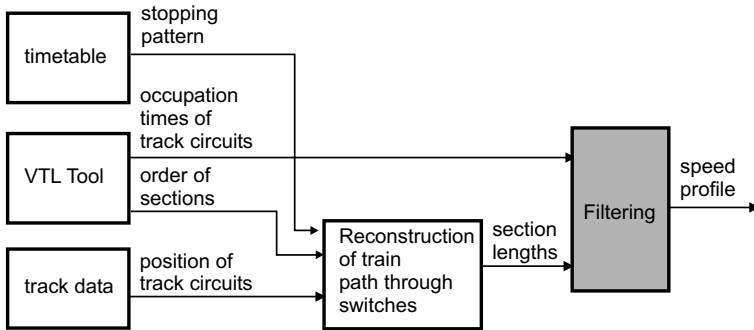


Figure 3: Implemented system components for speed reconstruction energy consumption estimation.

the assumed constant acceleration rate per section (cf. section 2.2). Those sections had to be matched with the stopping pattern of the train; in the case of long trains, the last section before the stop and the first section after the stop had to be identified.

3.3 Application for the RouteLint real-world tests

The proposed methodology was applied successfully for several hundred different trains riding on the corridor. The weighting factors have been chosen as $w_1 = 3D1000$ (per m), $w_2 = 3D1$ (per s) and $w_3 = 3D100$ (per m/s^2) respectively.

Because of the short section lengths of the examined line, the errors described in section 2.2 were only of minor importance: less than 5% of the examined trajectories contained the oscillations described there, mainly because of unplanned stops, but also because of very short sections (less than 100 m) passed with high speed (>120 km/h), that means occupation times of less than 4 secs (high influence of rounding error).

3.3.1 Passenger trains

Among the passenger trains, the Intercity trains which only stop in Rotterdam Centraal and Dordrecht are most likely to get into conflicts with the slower traffic on the route (regional passenger trains or freight trains). Fig. 4 shows the estimated trajectory of an IC travelling in southbound direction getting into a conflict situation with a single locomotive coming from the Kijfhoek.

The trajectory corresponds to the notes of the accompanying project ambassador. It says: "After a crossing train was visible on RouteLint, the driver slowed down the train." This happens at 35.5 km approximately: The train is slowed down and can approach the conflict area with a speed of 30-50 km/h. Not anticipating, it would probably have continued to go with maximal speed and later have come to a full stop, which would have caused more delay and a higher energy consumption.



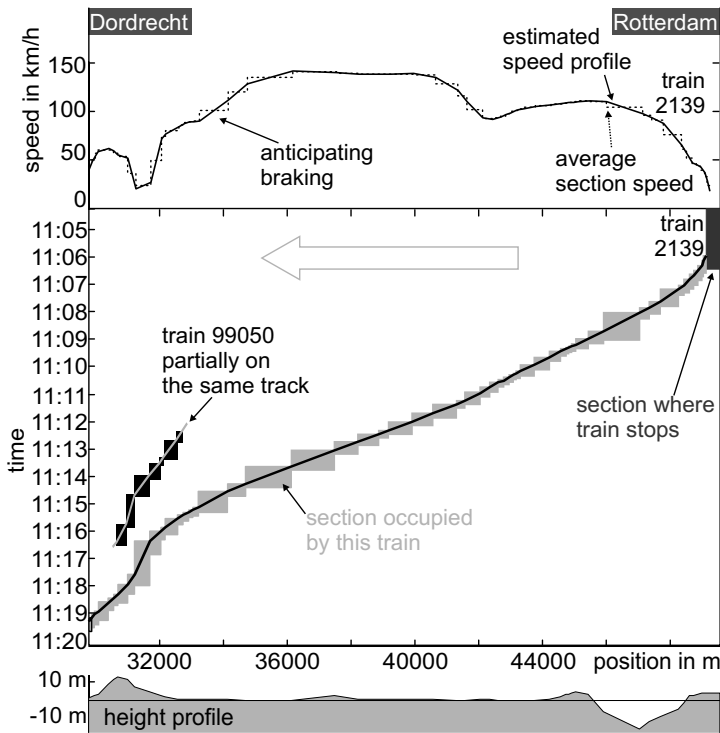


Figure 4: Intercity train with unplanned speed decrease.

3.3.2 Freight train

Fig. 5 shows trajectories of the local freight train 55007 for one train ride with and one without DIS. Due to the track geometry and bad signal visibility at some points of the line, the drivers without information on the signal states further on the line usually drive very conservatively in order not to have to brake the train from a high speed within a short distance. This can be seen from the trajectory without DIS, where the train comes through Dordrecht Station with a speed of around 40 km/h.

The DIS allows to look further ahead and to see that the route of the train is already clear: so with DIS, the driver did not have to ride that conservatively and could pass Dordrecht almost with the maximum allowed speed of 60 km/h saving as much as one minute travel time over the 3 km long section until the Kijfhoek freight terminal, which represents a saving of 20%.

The possibility of having faster freight trains can not be underestimated within the whole, very complex railway system: the infrastructure is occupied less with faster trains, less delays of freight trains lead to less conflicts with other trains, which itself allows reducing consecutive delays and therefore increasing service quality and reducing energy consumption at the same time.



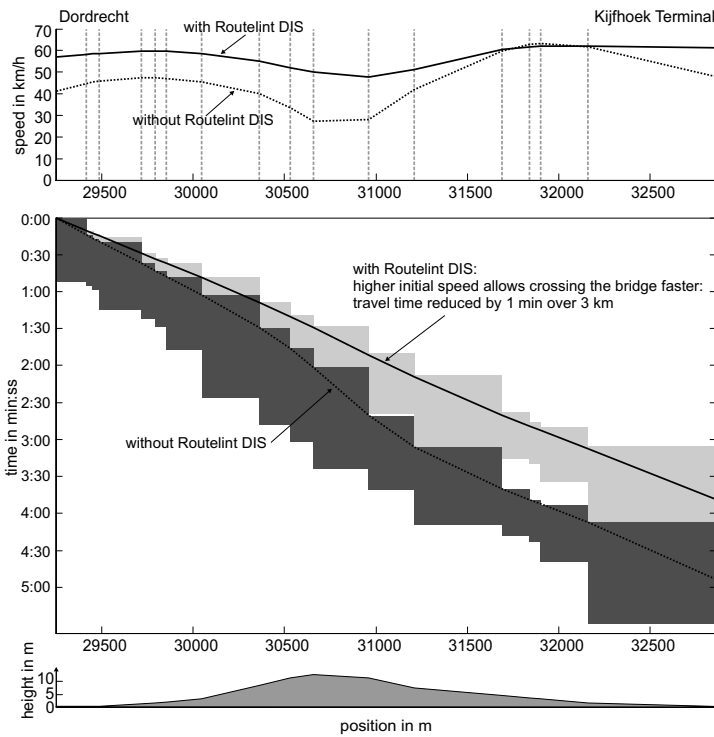


Figure 5: Freight train No. 55007 on 25-11-2005 (without DIS) and 01-12-2005 (with DIS).

This positive effect was also subjectively perceived by the freight train drivers who were most enthusiastic about the project.

3.4 Validation using energy meters

During the tests, five trains were travelling on the corridor which were equipped with energy meters: These devices delivered values for traction power every three seconds. This data has been used to demonstrate the correctness of the speed reconstruction approach. Fig. 6 shows one example of the reconstructed train speed over time for a fast regional train together with the measured traction power of the same train.

It can be seen, that both lines correspond well most of the time, however, there are some points (changes of regimes) where small time differences of less than 10 s can be observed (e.g. start of coasting between Rtd-Lombardijen and Rtd-Blaak).



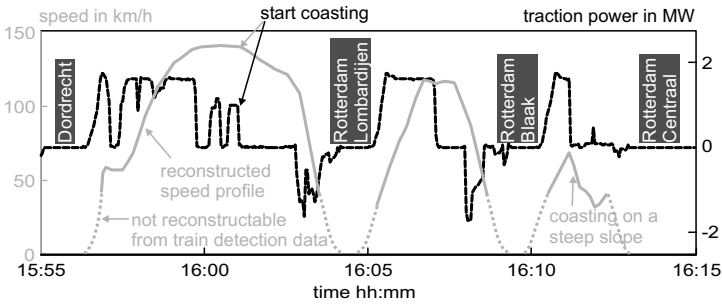


Figure 6: Fast regional train 2250 on 01-12-2005: Comparison of reconstructed speed profile with measured traction power.

4 Conclusions and outlook

The proposed algorithm delivered reasonable speed profiles, which corresponded perfectly with the observations of the accompanying project ambassadors in most of the cases.

For trains with high acceleration and braking rates, the given limitation of a single acceleration rate per section may cause oscillations which can be corrected with expert knowledge. Future research should focus on how to take that into account (varying number of acceleration changes per section, possibly under utilization of evolutionary algorithms as proposed for the computation of an unknown number of switching points for energy optimal train operation on sections with multiple gradients/ speed limits [4]).

Except for the reconstruction of train trajectories, the proposed methodology could also be used as an online application to compute current train speeds in a traffic control centre: This information should be valuable to the traffic controllers for train disposition in conflict situations and is available without having to install expensive measurement and communication equipment on board of every train.

References

- [1] Albrecht, T. & van Luipen, J., What role can a driver information system play in railway conflicts? *Paper submitted for the 10th IFAC Symposium Control in Transportation Systems*, Delft, The Netherlands, 2006.
- [2] Goverde, R.M.P., *Punctuality of Railway Operations and Timetable Stability Analysis*. Ph.D. thesis, Delft University of Technology, 2005.
- [3] Goverde, R.M.P. & I.A. Hansen, TNV-Prepare: Analysis of Dutch Railway Operations Based on Train detection Data, In: Allan, J., Hill, R.J., Brebbia, C.A., Sciutto, G., Sone, S. (eds.), *Computers in Railways VII*, pp. 779-788, WIT Press, Southampton, 2000.
- [4] Chang, C. & Sim, S., Optimising train movements through coast control using genetic algorithms. *IEE Proc.-Electr. Power Appl.*, **144**(1), pp. 65-73, 1997.



A decision support system for track maintenance

C. Meier-Hirmer^{1,2}, A. Senée¹, G. Riboulet¹, F. Sourget¹
& M. Roussignol²

¹ SNCF, Paris, France

² Marne-la-Vallée University, France

Abstract

For any railway infrastructure manager, track maintenance is one of the most important tasks to perform. This mission requires a regular follow-up of geometric track quality and monitoring of intervention execution. Among other tools, TIMON, a computer application using track geometry measurements and maintenance operation data, is currently in use in France. This application illustrates the evolution of track quality indicators.

We present a possible extension of this tool: a systematic analysis of the available data by applying a stochastic model for the track geometry deterioration is proposed. The use of environmental variables in order to identify the significance of their influence on track degradation is introduced.

The stochastic model has two advantages: on the one hand it enables one to take into account the intrinsic variability of the degradation phenomenon and on the other, significant environmental variables are included into the model. The work presented here aims to classify sections of the high-speed railway network with respect to similar degradation behaviour. Two different graphical representations that could ease decision-making are proposed.

Keywords: track degradation, stochastic process, maintenance scheduling, classification, decision support tool.

1 Introduction

In order to control track geometry, periodic levelling measurements are carried out by a measurement vehicle called Mauzin. The inspection step depends on the UIC group (indicator of the line importance). This study refers to high speed lines



where information is available every 45 days. Track geometry defects that can be detected by the Mauzin vehicle are presented in fig. 1. To each defect corresponds a specific measure.

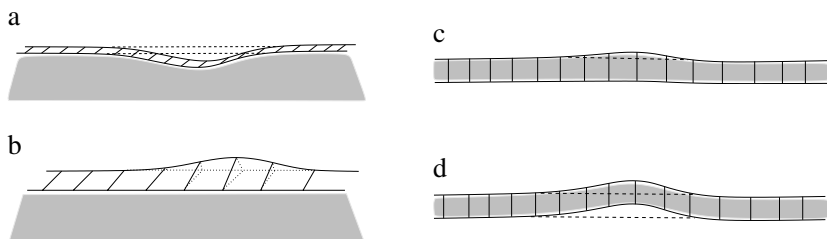


Figure 1: The different degradations of track geometry. a: longitudinal levelling, lateral view; b: transversal levelling, lateral view; c: gauge, top view; d: lining, top view. To simplify, only one selective defect is illustrated. The synthesised measures summarise the quality over 200 meters or one kilometre with respect to the type of degradation.

Using these measurements, indicators of track quality over 200 m or 1 km are calculated. The indicator over 1 km has been used for 30 years. Owing to new computer technology, an indicator over 200 m was developed at a later date. This new indicator is more detailed; it allows to better separate track sections containing switches and expansion joints from 'normal' sections. For example, using the older indicator the whole kilometre containing a switch has higher measurements. Using the 200 m indicator, only one of the five 200 m sections of the concerned kilometre has increased measurements. In order to maintain the history of the complete database, the 1 km indicator is still in use. There are five types of track quality: NL (longitudinal levelling), NT (transversal levelling), D (lining), Ec (gauge) and vEc (gauge variation). This article presents the use of statistical methods and their utility for track maintenance scheduling. For confidentiality reasons, the values of the significant environmental variables and the location of the different degradation classes are not given. The research presented in this article is based on the longitudinal levelling.

The article is built up as follows: section 2 describes an existing computer application TIMON (Traitement Informatique du Mauzin et des Operations de Nivellement: Computer processing of the Mauzin data and levelling maintenance) that is used at the SNCF for track degradation surveillance and for decision-making concerning major maintenance interventions. This tool displays information on the track geometry inspections as well as on maintenance operations carried out for each line.

Section 3 describes the data used in our study in addition to the NL indicator (3.1). This section also gives a short introduction to the statistical methods used to estimate the track degradation rate and to generate classes by cutting the high



speed line into fractions having similar degradation (3.2). We apply these methods to real data and present the results (3.3).

Section 4 shows two possible types of diagrams that can be used to include the results of the analysis presented in section 3.3 into the existing decision-making tool for track maintenance TIMON.

The article ends with a discussion on proposed improvements that could ease decision-making.

2 The existing decision-making tool: TIMON

In order to display and to treat data coming from the Mauzin measurement vehicle, a computer application named TIMON was developed by the infrastructure information system department. For other decision-aiding applications see Le Bihan [1]. The synthesized indicators mentioned in section 1 are calculated after every inspection and are included in the TIMON database. This database contains, among other things, all synthesised Mauzin indicators ever measured on high speed lines. The inspections on these lines are usually carried out every 45 days. We are particularly interested in the NL indicator (*cf.* explanation in section 1) as it is often used for assessing the necessity of major interventions.

In the top part of fig. 2 we can see different interventions carried out associated with the length of the concerned section. At the upper right-hand side switches or bridges are indicated. For example, the selected section, which is displayed in fig. 2 includes a bridge. fig. 2 shows also the development of the NL indicator. We have chosen the 1 km section, but the same information is available for the 200 m sections. It can be seen from the graph that the tamping carried out during the first trimester 2004 reduced the value of the NL indicator. A special feature of this type of display is its aid in predicting the exceeding of thresholds. It is possible to select NL measurements and to conduct a linear regression on these points. The date at which the threshold is likely to be exceeded, is also calculated.



Figure 2: The TIMON application: NL development.



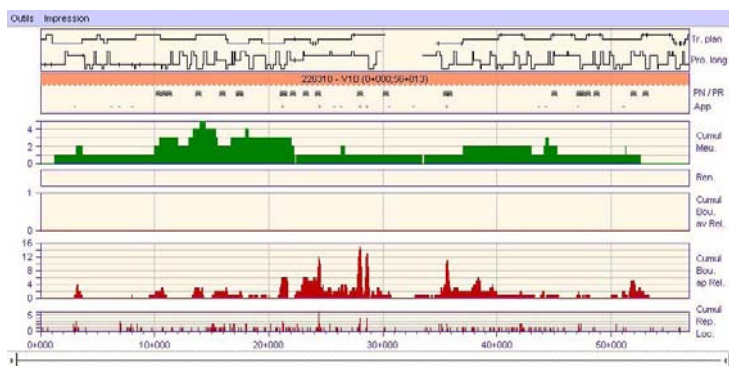


Figure 3: The TIMON application: cumulated interventions.

The work presented in this article aims to ameliorate this calculation by including the uncertainty of the track development. We are therefore using a stochastic process for modelling. Our aim is not only to display the track characteristics but to include them directly in the prediction of track development. TIMON is able to display all types of track interventions (tamping, grinding, track lifting, ballast and rail replacement). Information on track characteristics, like curves or switches, is also given.

Fig. 3 shows one of the possible screens: the cumulated intervention display. The number of grinding and tamping interventions is calculated. With respect to tampings, it is even possible to separate the number of tampings before and after a track lifting or ballast replacement. As this screen gives supplementary information about track characteristics, it is possible to verify if there is a correlation between these environmental variables and the number of interventions.

It is also possible to compare the measurements of selected inspections for a whole line or a part of a line. Other features include the search for extreme synthesised indicator values and the computation of the mean degradation over a selected section including or excluding bridges, tunnels, switches and expansion joints.

The display of real Mauzin measurements (in addition to the synthesised indicators) and the acceleration measurements is under development. A prototype, which allows in addition a combined analysis of these indicators, is tested at the moment.

3 Track degradation study

3.1 Available data

The statistical analyses are based on track geometry measurements of French high speed lines. We use the synthesised indicator NL for this study. Data are available for every high speed line since its opening. Inspections are usually carried out



every 45 days. In order to explain the degradation difference between different track sections, we introduce environmental variables (covariates). The following information is incorporated into the study:

- ballast: type, height of sub-ballast;
- ballast renewal;
- maximum speed;
- UIC group;
- curves and weighted average of curvature;
- cant and cant deficiency;
- ascending and descending gradients;
- combination of curves and gradients;
- tonnage in 2004 and cumulated tonnage since the last ballast renewal;
- climate: temperature, precipitation and frost;
- sleepers: type and year of fabrication;
- age of the line;
- rail: type and treatment;
- tunnels and bridges;
- expansion joints and switches.

The climate is only available for every administrative district. These two covariates are therefore highly correlated. The information about the climate may be too coarse to give coherent results.

The NL indicator and the environmental variables are merged into two databases; one with a 200 m, the other one with a 1 km resolution. There is no specific treatment for the synthesised indicator, as the information is already available for every 200 m and 1 km. In order to get the same splitting for the covariates, the following rule is used: the predominant value with respect to the length is attached to every 200 m or 1 km section. Tunnels and bridges are entered if they are longer than 200 m. Expansion joints and switches are considered if they exist on the concerned section. Altogether we have two databases with 31 categorical and 21 continuous covariates.

3.2 Methods

3.2.1 The Gamma process

The goal is to propose an estimate of the track degradation rate incorporating the uncertainty about track evolution. The estimate should be based on the NL indicator, not just on adjacent measurements between two interventions. The TIMON application can only treat the latter analysis (*cf.* 2). We use the stochastic Gamma process for modelling. The Gamma process $X_{t,t \geq 0}$ has the following properties:

- X_t has independent increments,
- for $s \geq 0$ and $t \geq 0$, $X_{t+s} - X_t$ has a Gamma distribution with the parameters αs and β . The density is given by

$$f_{X_{t+s}-X_t}(x) = \frac{1}{\Gamma(\alpha s)} \beta^{\alpha s} x^{\alpha s-1} e^{-\beta x} \mathbb{I}_{\{x \geq 0\}}.$$



where $\alpha > 0, \beta > 0$. Therefore the Gamma process has a constant mean degradation rate ($\frac{\alpha}{\beta}$) and a constant degradation variance rate ($\frac{\alpha}{\beta^2}$).

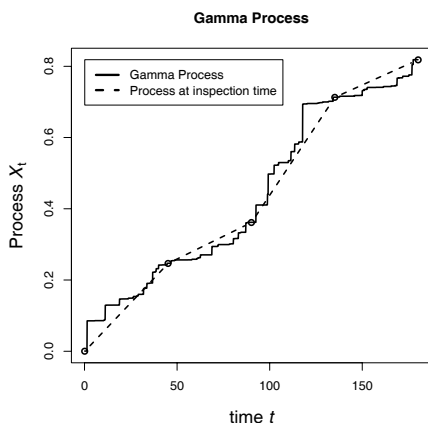


Figure 4: Trajectory of the Gamma process and the information known at inspection times.

Fig. 4 shows the trajectories of the Gamma process and explains the information loss due to the periodic inspections. Only continuous surveillance could provide an observation of the exact trajectory. The slope of the dashed line is approximately $\frac{\alpha}{\beta}$. The variation of the solid line from this mean degradation is related to the quantity $\frac{\alpha}{\beta^2}$. The properties of the Gamma process comply with SNCF expert judgement and it was verified that it gives satisfactory results. Simulations of the process development were performed and compared to real data.

Real NL indicator data is used to estimate α and β for every 200 m or 1 km unit of all French high speed lines. Values ($\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$) are then calculated.

3.2.2 Classification method

The method used to explain the track degradation based on environmental variables is called CART (Classification and regression trees). This method was introduced by Breiman *et al.* [2]. A feasibility study was conducted by Senée [3] and gave reasonable results. We decided, therefore, to retain the CART method.

Regression trees can be used to predict the value of a continuous output variable based on one or more input variables. The input variables can be a mixture of continuous and categorical variables. The “regression tree is built through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions, and then slitting it up further on each of the branches” [4]. The process works as follows:

1. The whole data is located in one class.
2. For every variable and for every possible split, the algorithm calculates the sum of the squared deviations, which is an indicator of the homogeneity of



the partitioning. The partition minimising this criterion is kept.

3. The algorithm repeats point number two, above, with the two new branches.

This process continues until one class is created for every observation. “However, this would not make much sense since one would likely end up with a tree structure that is as complex and ‘tedious’ as the original data” [5]. In our study we use two stop criteria:

- the complexity level of the tree is predefined regarding the length of the concerned railway line,
- splitting is not realised if a partition regroups less than five observations for the 1 km sections and ten for the 200 m sections.

The use of regression trees has the following advantages: “Tree methods are nonparametric and nonlinear. The final results of using tree methods for regression can be summarised in a series of logical ‘if-then’ conditions (tree nodes). Therefore, there is no implicit assumption that the underlying relationships between the predictor variables and the dependent variable are linear, follow some specific non-linear link function, or that they are even monotonic in nature” [5].

The following section shows the results of this method applied to our data. “In most cases, the interpretation of results summarised in a tree is very simple” [5]. In comparison with other methods, tree methods “can often yield a much simpler ‘model’ for explaining why observations are classified or predicted in a particular manner” [5].

3.3 Results

The methods described in section 3 are used to estimate the mean degradation rate $\frac{\alpha}{\beta}$ and the degradation variance rate $\frac{\alpha}{\beta^2}$ every 200 m or 1 km, respectively. The computation is done using the R software [6]. Classification trees are constructed for the two objective values ($\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$) and for the 200 m and 1 km sections. We thus dispose of four trees per line. Trees based on the data of all French high speed lines are also constructed. The computation was done using SAS[®] Enterprise Miner[™] [7]. For confidentiality reasons we cannot give the values of the separating variables.

Fig. 5 (a) shows the results of the mean degradation rate for the 1 km resolution. The variable TOTTON has an important impact on the degradation rate. TOTTON is the total tonnage of the line since the opening. Another variable that separates well the sections in two classes is the variable expansion joints.

Fig. 5 (b) presents the results of the mean degradation rate for the 200 m resolution. Again we find the covariate TOTTON. Another covariate which divides the line sections into two groups is CRV+GRAD. This covariate indicates where there are significant curves and at the same time significant gradients. Both trees divide the line into three degradation groups. In the following section (4) we propose graphical representations of the three classes.

The trees per line give interesting results. It is possible to identify covariates that were judged so far not to be the most influencing ones, such as the combination of curves and gradients. But even in the case where the trees confirm expert



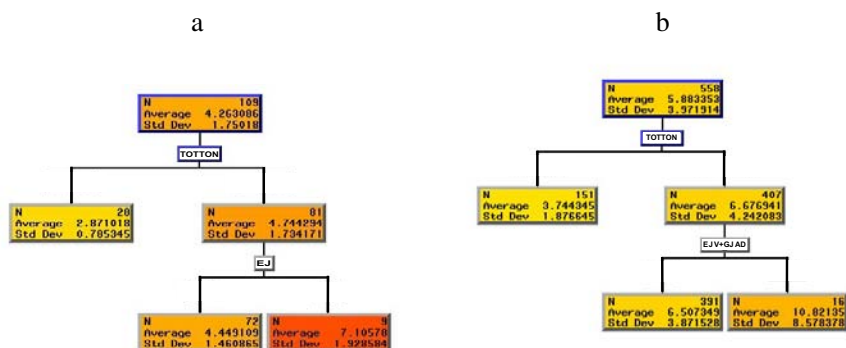


Figure 5: Classification tree for a high speed line; (a) 1 km resolution (b) 200 m resolution.

knowledge (influence of speed and tonnage) they are able to rearrange these covariates according to their importance for every specific line.

The global trees, based on the data of all high speed lines, bring out covariates which are very different between the lines and whose importance was emphasised by the experts already before the analysis (*e.g.* tonnage and age). The trees for every single line, however, show some interesting results. The covariates specific to every line will be investigated by track maintenance experts. There are some lines where the 200 m and the 1 km trees gave similar results like in the case we presented here; for others, the results are very different. This can be explained by the enhanced accuracy of the 200 m indicator and by the slight difference in the CART algorithm (minimal class size of five or ten observations; *cf.* 3.2.2). A similar phenomenon exists for the mean degradation rate trees and the degradation variance trees.

4 Improvements of the existing tool

The following pictures are screen shots coming from a graphical display application for infrastructure data developed at the Innovation and Research Department of the SNCF. In order to propose improvements of the maintenance surveillance tool TIMON, we added our results manually.

The first proposal is to use a map of France to visualise the different degradation classes (*cf.* fig. 6). This design takes account of the mental model principles formulated by Wickens *et al.* [8] for interface design. One of the principles is that the design display should look like the object that it is trying to illustrate. Using this type of display, it is easy to locate the zones having good or bad track degradation behaviour.

Another representation is shown in fig. 7. The advantage of this type of display is the fact that it is based on the long-existing track layout plans. Maintenance experts



are used to this type of illustration, already incorporated in the existing application TIMON. Some of the covariates used for the construction of the trees are available on the track layout plans (*e.g.* tunnels and bridges). It is therefore possible to better understand the partitioning created by the algorithm. From an ergonomic point of view, this type of representation respects the memory principle. Information similar to the information delivered by the existing tool, is displayed in a similar way. Thus, the different displays are consistent in its information projection.

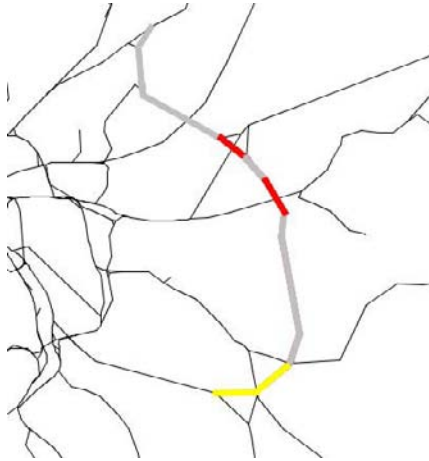


Figure 6: One possible presentation of the track degradation classification results is the plot on a map of France. Displaying the results in this way facilitates the geographical localisation of line parts having characteristics leading to good or bad geometric behaviour. Every colour corresponds to a degradation class.

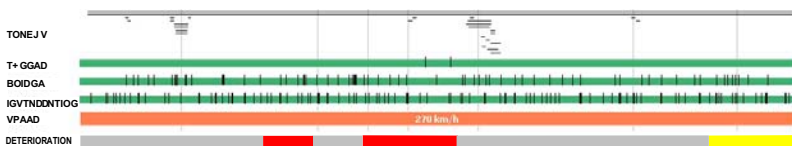


Figure 7: A second presentation of the track degradation classification results. The line is represented by a vertical line. Displaying the results in this way has the advantage that additional covariate information, such as height of the track or the presence of bridges, can be indicated together with the degradation class.

5 Conclusion

Our study uses environmental variables in order to explain track degradation and to create classes of homogeneous track quality. The chosen CART method seems appropriate for the task to accomplish. The classification trees show the importance of covariates so far not used for maintenance scheduling and sort the covariates known by the experts according to their influence for every line. Based on these results, we propose improvements of an existing track maintenance surveillance tool that could ease major operation allocation.

There are two major perspectives for the study. In the first place, we want to include even more covariates that could be important for the track degradation. A larger inspection history and more precise covariate information could enhance the analysis. For example, the climate data are too coarse. Efforts are currently being made to obtain more detailed information. In the second place we want to insert the results into the TIMON application. One of the graphical representation of the results has to be chosen and the corresponding code has to be developed. The results of the trees will be carefully analysed before incorporating them into the decision-making tool and before using them for maintenance scheduling.

Acknowledgements

We want to thank V. Kulnieks for his helpful advice concerning the English language and C. Horsch for proof-reading this article. We also thank S. Lebourgeois and L. Centonze from the LINCOLN society, who did the required computation for the tree construction. We thank J.Heras from the Infrastructure Information Service for spending time explaining the TIMON application.

References

- [1] Le Bihan, A., Évolutions de la maintenance de la voie à la SNCF. *Congrès International Voie*, 1997.
- [2] Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J., *Classification and regression trees*. Chapman & Hall, 1984.
- [3] Senée, A., *Analyse des indicateurs des défauts de géométrie de la voie ferroviaire*. Master's thesis, ENSAI (École Nationale de Statistique et de l'Analyse de l'Information), 2005.
- [4] Resampling Stats, http://www.resample.com/xlminer/help/rtree/rtree_intro.htm
- [5] Statistica, <http://www.statsoft.com/textbook/stcart.html>
- [6] R project, <http://www.r-project.org>
- [7] SAS Enterprise Miner, <http://www.sas.com/technologies/analytics/datamining/miner>
- [8] Wickens, C.D., Gordon, S.E. & Liu, Y., *Introduction to Human Factors Engineering*. Addison-Wesley-Longman, 1997.



The new Shinkansen rescheduling system for drivers and crew

H. Shimizu¹, H. Tanabe², S. Honda² & K. Yasura²

¹*JR East Japan Information Systems Company, Japan*

²*East Japan Railway Company, Japan*

Abstract

We are now undertaking development of the new rescheduling plan system for drivers and crew, to synchronize with the forecast train plan. The East Japan Railway Company has five Bullet Train (Shinkansen) lines operating about 800 trains per day, dispatching nearly 250 drivers and 350 crew per day to make the trains operational. To keep the operation plan and management of Shinkansen in order, we have the system named COSMOS (Computerized Safety Maintenance and Operation systems of Shinkansen). The drivers and crew work plans are made with the drivers and crew rescheduling system, one of the subsystems of the COSMOS transportation planning system. Each driver and crew works on their train according to the work plan normally decided on beforehand. However, sometimes trains would be unable to operate on schedule when weather is bad or car troubles occur. In such cases, the work plan must be redesigned. Should the change affect two or more crews' the redesigned plan should only change the plan of one crew member. It is very difficult to make this plan change because the delay of the train changes minute by minute. To change the crews' work plans adequately, we have developed a new system. This system displays the contradictions of the change plans of the crew if the train delay happens. Work planners make the change plan to solve these contradictions.

The changing of jobs can be selected for the most suitable options, accurately and quickly, based on the computer supporting systems. We reduce the delay time of trains by utilizing this system and it will improve our services for customers.

Keywords: train scheduling, driver scheduling, crew scheduling, forecast, real-time rescheduling.



1 Introduction

The East Japan Railway Company has the latest computer system that holds the plan and the management of an elaborate train schedule of SHINKANSEN. This system enables a safe and accurate train operation.

This system is called “Computerized Safety Maintenance and Operation Systems for SHINKANSEN” – “COSMOS”. COSMOS has eight sub-systems – the “Transportation plan system” called “TPS”, the “Operation control system” called “OCS”, the Railway yard work management system, the Maintenance works control system”, the Rolling stock control system”, the Facility control system”, the Facility monitoring system” and the “Electric power control system”. In COSMOS, these eight sub-systems are cooperating elaborately.

We have already been using COSMOS for 11 years since 1995. We are now advancing the redevelopment of TPS and OCS. It aims at a further function improvement, and the development of the new feature that uses a new technology – for example, the Train rescheduling system using constraint programming, etc. – is included in this development.

In this thesis, “Rescheduling for driver and crew, using forecast train scheduling data” of the new features is described.

1.1 About forecast train scheduling data

In the OCS, the forecast train scheduling data is made with the system so far. This data is a calculation of the scheduled departure and arrival times at each train station and the forecasted arrival and departure times according to actual events. The forecast train scheduling data, called “FTS”, periodically displays the information once a minute in this system. When the train is delayed, this system calculates the delay time of the departure and arrival at each station.

1.2 About rescheduling for drivers and crew

TPS includes two scheduling systems for drivers and crew. The first one is a function to perform planning and management of the driver and crew scheduling several months in advance. Another function, as driver and crew scheduling is planned beforehand, reschedules when the train is delayed. This function is called “Rescheduling for driver and crew”. Hereafter, it is called “RDC”.

2 Comparison and new feature with current functions

2.1 Current problems

The current RDC has some problems. In this paragraph, current problems are enumerated and each solution for them is described in the next system.

There are the following problems.

- (a) The plan of drivers and crew cannot be rescheduled to synchronize with the delay of trains.



- (b) The plan of drivers and crew cannot be rescheduled to synchronize with the rescheduling of trains.
- (c) If no adjustment happens, within a short time the current system fails.

We introduce the case, (b) that can actually happen as follows.

There is a plan that the driver changes from train “1” to train “2” at station A. Figure 1 shows this. At this time, using the track of station “A”, the train is changed from track “2” to “1” in the train plan as shown in Figure 2.

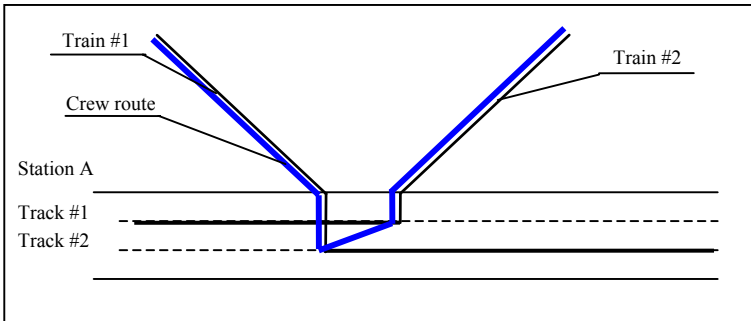


Figure 1: Before track changing.

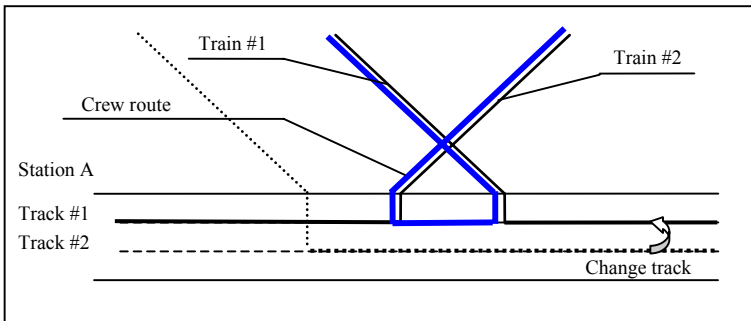


Figure 2: After track changing.

As a result, no adjustment is generated in the train plan and the driver plan.

In a word, train “1” cannot arrive at station “A” if train “2” doesn’t leave. On the other hand, if train 1 doesn’t arrive, train 2 cannot depart, because the same driver works on both trains.

In the current system, such an event might not turn out satisfactorily.

The plan has a big influence as the train is delayed when such an event happens.



In the new system, the mechanism that no adjustment can be understood at the early stage is necessary to prevent such an event beforehand.

2.2 For the problem solution

We assumed that we added the following function to new RDC.

- (a') FTS is reflected in real time to RDC.
- (b') The plan change of trains is reflected in real time to RDC.
- (c') Enables us to understand no adjustment of the plan that happens previously in several hours.

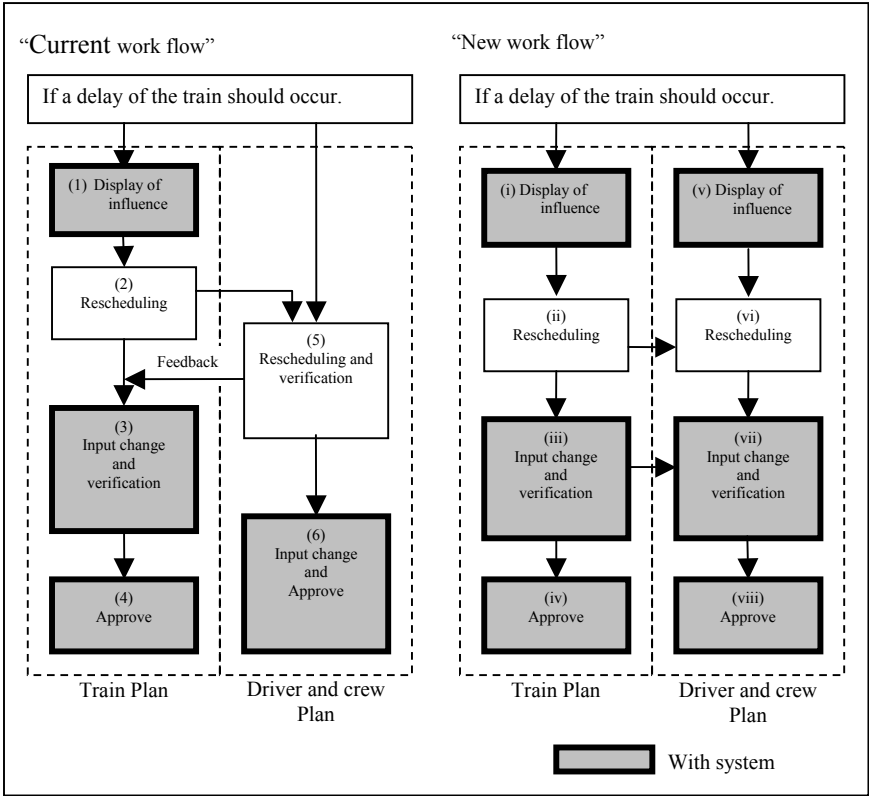


Figure 3: Current and new work flow.

Figure 3 shows current and new workflow for drivers and crew rescheduling. (a') corresponds to (v) of Figure 3. It is shown to be able to display the delay of the train influences on planning with the system.

(b') corresponds to (vii). A current system takes the change in the train rescheduling plan like (5) outside the system. A new system takes the change like (vii) inside the system.



(c') is achieved as a whole by cooperate with the train rescheduling and the drivers and crew rescheduling.

3 New functions of RDC

This chapter describes the new feature of RDC.

3.1 Reflection of FTS in RDC

Each minute or every instance when the plan is changed, the system makes FTS. Each time, this FTS is stored in the database. RDC periodically gets FTS from the database. FTS is reflected in the diagram screen and the horizontal bar screen in RDC.

Figure 4 shows the example of displaying the screen when the train is delayed.

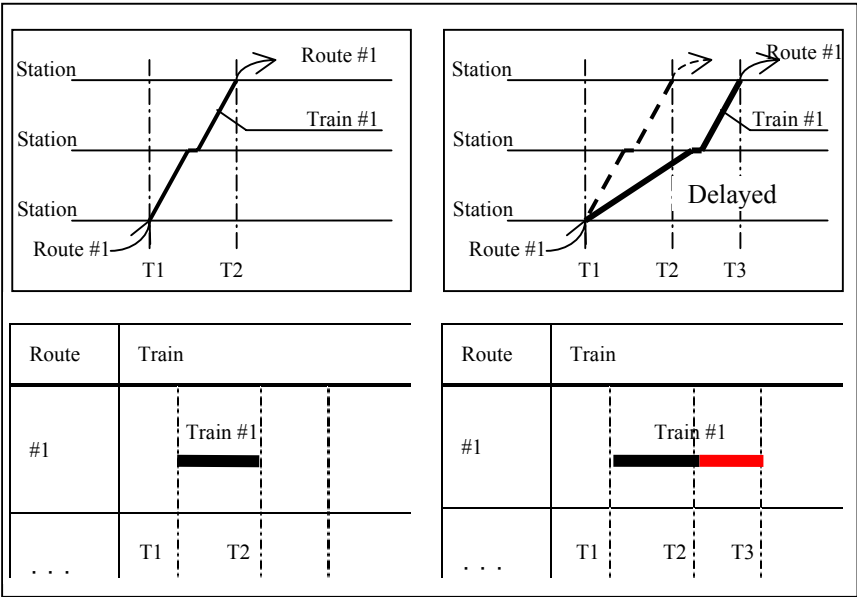


Figure 4: Example of displaying the screen in RDC.

3.2 Display of change recommendation point of plan by FTS in RDC

This chapter explains three functions of RDC. It is the one to display both of these three functions plan change recommendation points.

3.2.1 Attention of change of drivers and crew plan

This attention means that the driver or crew will be late for the train that works next, because the train on which they are working is delayed. In this case, the attention mark is displayed on the screen.

There is a plan that a certain driver or crew works with train “1” and train “2”. He changes it from train “1” to train “2” at station “C”. At this time, the condition of displaying this attention mark is shown by the following expressions.

$$T_a(1, C) + N(C) > T_d(2, C) \text{ ----- formulate 1}$$

$T_d(r, s)$: Forecast departure time
 $T_a(r, s)$: Forecast arrival time
 $N(s)$: The change time required
 R : Train number
 S : Station

Figure 5 shows the example of displaying the screen.

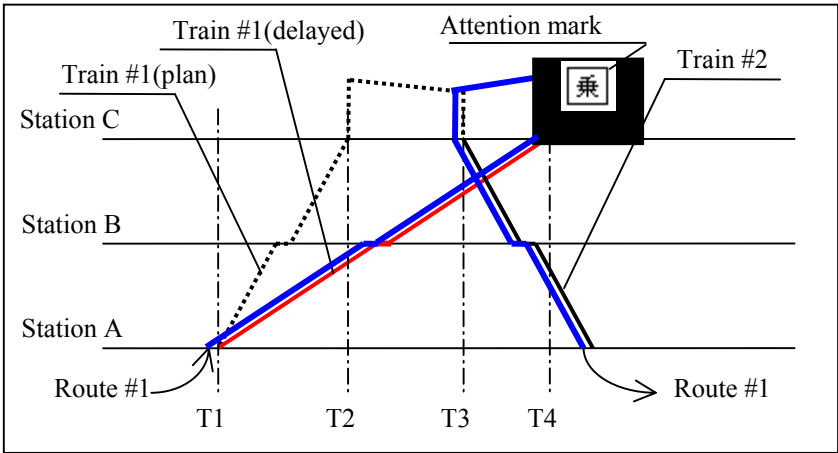


Figure 5: Attention mark of change of drivers and crew plan.

3.2.2 Attention of the train assigned no driver or crew

When one train is newly added in the train rescheduling, driver or crew has not been allotted this new train in RDC yet.

At this time, on the screen, it is displayed that the color of the line of this new train is different from usual. Moreover, the round mark painted out in white is displayed at the first station of the section that not is even if the driver or crew allocates it. Figure 6 shows this.

3.2.3 Attention of difference between station getting on and getting off

When one train becomes driving discontinuance among trains by which a driver or crew is scheduled to work, the station that he gets off with the train ahead of that and the station gets on with the train afterwards are different.

At this time, the attention mark is displayed on the screen. And a line is displayed between the station of getting off and the station of getting on.



In Figure 3, it is shown that the station where driver or crew gets off train “1” and the station where he takes train “3” are different by discontinuance of train “2”.

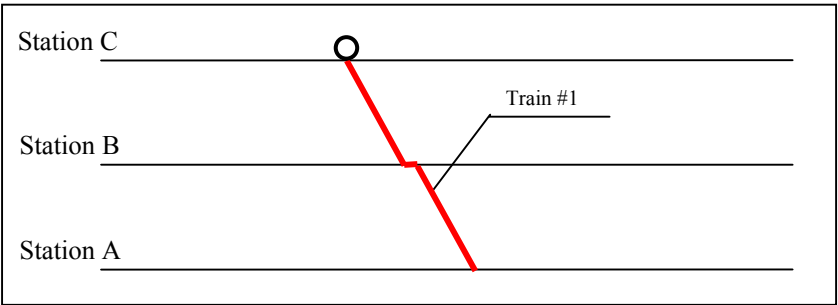


Figure 6: Attention of the train assigned no driver or crew.

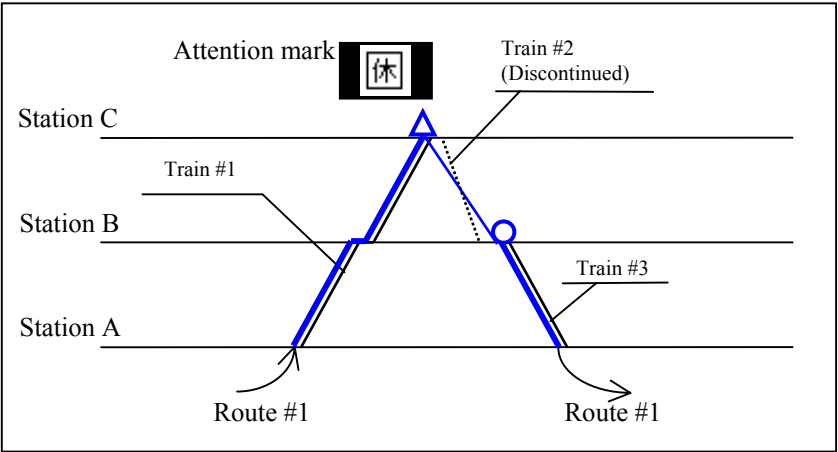


Figure 7: Attention of difference between station getting on and getting off.

4 Conclusions

In this thesis, the cooperation of FTS and RDC has been described. We think that RDC becomes more strategic and more efficient by cooperating with FTS.

In the railway transportation, RDC is very important. The Shinkansen transportation doesn't work out even if there is a correspondence only in the train plan. It works out with a correspondence in both the train plan and the drivers and crew plan.

We can understand no adjustment of the plan that happens previously in several hours with new RDC.

For example, we can understand for several hours or more though we were able to understand driver's or crew's lack only immediately before the current.



Then, we can beforehand let a preliminary driver go to the departure station of the train allocated nobody, and can let him get on the train.

From the above-mentioned, we think that RDC becomes more flexible now, and the Shinkansen transportation becomes stabler.



A Decision Support System for railway timetabling (MOM): the Spanish case

F. Barber¹, P. Tormos², A. Lova², L. Ingolotti¹, M. A. Salido¹
& M. Abril¹

¹*DSIC, Universidad Politécnica de Valencia, Spain*

²*DEIOAC, Universidad Politécnica de Valencia, Spain*

Abstract

Train timetabling is a difficult and time-consuming task, given the complexity and number of constraints to be taken into account. A feasible train timetable should specify the departure and arrival time of each train to each location of its journey so that the line capacity and other operational constraints are met. Traditionally, train timetables have been generated manually by drawing trains on the time-distance diagram, where train schedules are manually adjusted so that all constraints are fulfilled. This process can take a long time and it usually stops once a feasible timetable is found. The difficulty of the process increases in complex and high-loaded networks and the resulting timetabling may be far from optimal. In this work, we present MOM: a friendly, flexible, computer-based Decision Support System (DSS) whose main goal is to support the railway planner to efficiently obtain optimized train timetables. The system generates feasible and high quality timetables by considering the set of constraints of the problem and optimization criteria related to the use of railway infrastructures and operator requirements. The railway planner can obtain several solutions automatically, by modifying certain parameters (departure interval, frequencies, among others). Hence, the user can analyze them and take decisions about conditions related to railway traffic and infrastructure. Furthermore, an important feature of the DSS is the consideration of hard and soft constraints which makes the search process more flexible and reaches better solutions. The developed DSS offers assistance in train scheduling, conclusions about the maximum capacity of the network, identifying bottlenecks, determining the consequences of changes, providing support in the resolution of incidents, and alternative planning in on-line scheduling for real traffic control. This DSS software is being successfully used by the Spanish Manager of Railway Infrastructure (ADIF).

Keywords: Decision Support System, train timetabling, railway capacity, on-line scheduling.



1 Introduction

European railways are undergoing a period of massive reforms that include the separation of infrastructure management from operators, which involves the open access to new operators competing for the use of the available infrastructure. This new scenario requires an improvement in the use of existing railway infrastructures, increased efficiency and the consideration of a more customer-oriented view looking for increasing the quality of the services they offer. The Spanish Manager of Railway Infrastructure (ADIF) is also involved in this process and is continuing the direction started some years ago, which is focused on the improvement of the use of the network capacity and the increase in quality of the services offered to customers. The Spanish railway network consists of 13,177 km of line that supports a heterogeneous traffic with around 5,000 daily trains, travelling a total of 494,230 km per day. The topology of the Spanish network is radial and is managed taking into account the scheduling of lines with few changeovers. Since is to make the best use of the available resources, one of the specific goals of ADIF is the automatic generation of train timetables that fulfil the existing constraints and optimize several quality criteria.

The MOM project is a long-term collaboration between the Polytechnic University of Valencia and ADIF. The aim of the project is the development of a Decision Support System that offers assistance in train scheduling, obtaining optimized timetables, conclusions about the maximum capacity of the network, identifying bottlenecks, determining the consequences of changes, providing support in the resolution of incidents, providing alternative planning, and real traffic control, etc.

There are several, useful software systems for railway planning management: RAILSYS, DONS, PETER, ROMAN, VIRIATO, CAPRES, FASTA, AFAIG, DEMIURGE, SISYFE, RAILCAP, OPENTRACK, OPENTIMETABLE, VISION, CMS, TPS and MULTIRAIL among others [2, 4]. All these systems have Infrastructure Management, Evaluation, and Simulation Management modules. Some of them have capabilities of on-line scheduling, timetable optimization, and capacity analysis. All these modules are integrated in MOM with different levels of development. Therefore, MOM can be considered as one of the most complete DSS software systems for railway timetabling, since it can support managers of railway infrastructure, in the difficult decision-making process they face.

2 Railway problem description

Given a railway line which may have both single and double-track sections, the railway timetabling problem consists in adding new trains (periodic or non periodic) to the line in such a way that the timetable assigned verifies the existing constraints and optimizes a multi-objective function. The railway line may be occupied by other trains whose timetables cannot be changed, and the new trains may belong to different operators with different priorities. The stations to be visited by each train may also be different from each other. The



objective of the system is to obtain a feasible and optimized timetable taking into account: (i) the railway infrastructure topology, (ii) user requirements (parameters of trains to be scheduled), (iii) traffic safety rules, (iv) traffic previously scheduled on the same railway network, and (v) optimization criteria. Table 1 shows the main constraints that are taken into account by MOM.

Table 1: Constraints considered by MOM.

Railway Timetable Constraints	Type
Time Interval for the Departure of the First Station	Operator
Time Interval for the Arrival to the Final Station	Operator
Maximum Delay allowed	Operator
Frequency (for a set of periodic trains)	Operator
Commercial Stops in Stations	Operator
Journey Times (for each track section)	Operator
Avoidance of Crossings (in one-way track sections)	Traffic
Avoidance of Overtaking on track sections	Traffic
Delay for a non specified commercial stop	Traffic
Reception/Expedition times in crossing/overtaking operations.	Traffic
Avoidance of Simultaneous Departures in stations	Traffic
Finite track capacity of stations for traffic operations	Infrastructure
Station Closing Times for traffic operation	Infrastructure
Headway Times	Infrastructure

3 General Architecture of MOM

We have developed a flexible software system whose main goal is to assist the railway planner to design high-quality train timetables with a reasonable computational time. MOM generates feasible and high quality timetables by considering both the set of constraints of Table 1 and the multi-objective function specified. Moreover, the railway planner can also modify different parameters (departure interval, maximum delay, closing time, number of tracks, etc.) to adjust user/infrastructure requirements and obtain different solutions, analyze them, and take decisions about both railway traffic conditions and the available railway infrastructure.

Figure 1 shows the basic architecture of MOM. We have structured MOM in two main components: the interface and the optimizer. The interface allows the user to configure the scenario on which the optimization will be carried out; and the optimizer is responsible for providing the planner with a feasible, high-quality timetable according to the constraints specified and the user requirements.

The interface module acquires the required railway infrastructure data directly from standard databases. The optimizer implements advanced Constraint Satisfaction and Optimization techniques to find feasible and high-quality timetables. A detailed description of the scheduling methods integrated in MOM



is available in [1, 3, 5]. Once the first timetable is obtained, it is available to the user in a graphical or textual format, together with a detailed report about its main characteristics. Afterwards, MOM can obtain other successive and alternative timetables.

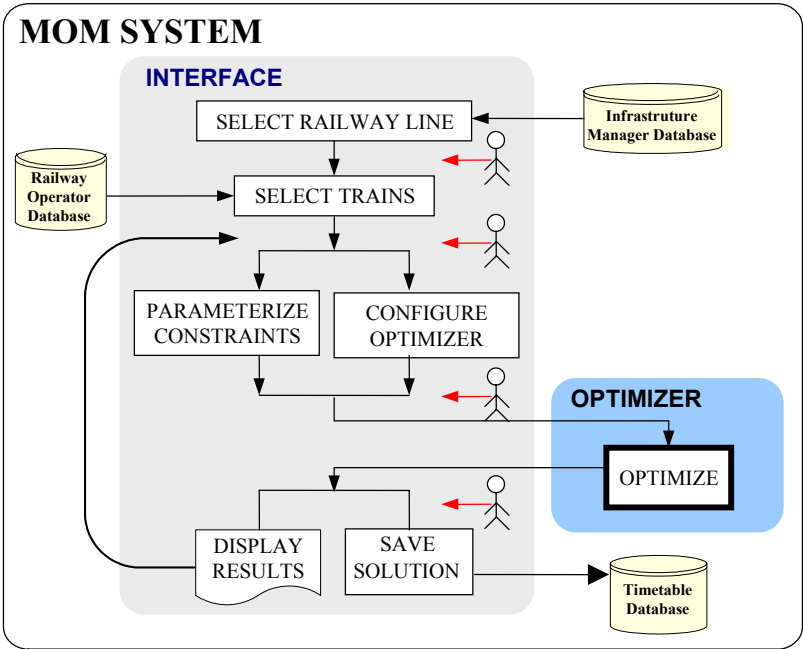


Figure 1: Components of MOM Interface.

In the interface module of MOM, the user can establish parameters corresponding to constraints, railway infrastructure, or trains. The main interface functions are described in the following points.

3.1 Railway line and train selection

The user selects the railway line to be used and the trains that will be taken into account by the optimizer. Some of these trains can be considered as trains in circulation. Therefore, their timetables are fixed and will not have to be modified during the solving process. The rest of the trains will make up the List of Trains to be Optimized (LTO), that is, the trains that need to be made compatible with the fixed trains over the given railway infrastructure.

3.2 Journey, train and traffic constraints

The main journey, train, and traffic constraints that the user can parameterize are shown in Table 1: the time interval for the departure of the first station and/or arrival to the final station for each train, the running time for each track section

corresponding to the journey of each train in the LTO, maximum delay allowed for each train, the headway time between consecutive trains in the same direction, the reception and expedition times for crossing or overtaking operations, minimal time interval between consecutive departures in stations, the finite track capacity of stations, the closing time in stations for traffic operations, the frequency for periodic trains, the time intervals for infrastructure maintenance, the selection of locations for technical stops of some trains, and the delay for a non specified commercial stop due to unexpected acceleration and braking times. The main interface of MOM is shown in Figure 2.

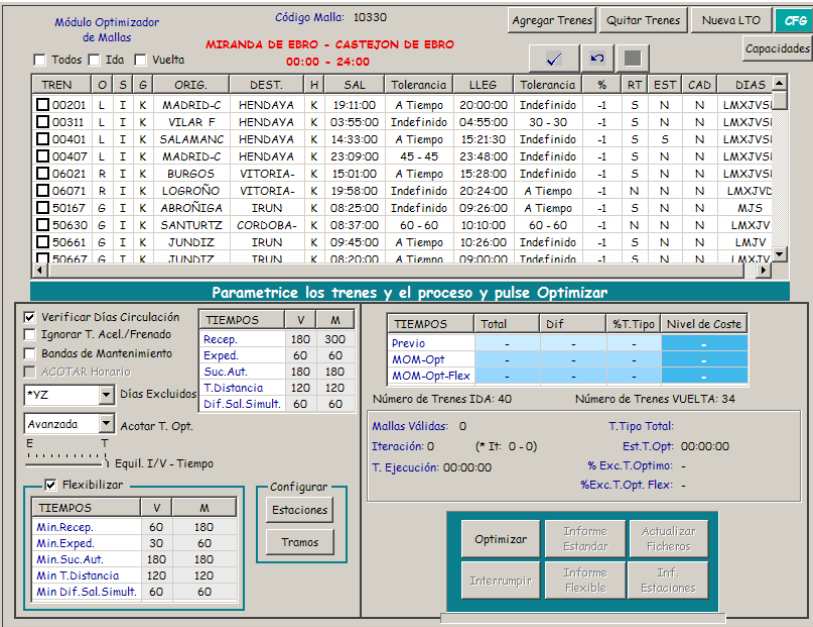


Figure 2: Main interface of MOM.

3.3 Optimizer configuration

The optimizer module uses a multi-objective function to assess the quality of feasible timetables. This function is composed by several weighted criteria: the average delay for each train, the number of traffic operations, the balance of the delay between upward and downward trains, etc. The weight assigned to each criterion is determined by user. According to the weight given to each criterion, the process searches for the solution that best fulfils the considered objective function. Furthermore, the user can bound the computational burden allowed for optimization. Hence, the obtained timetable will always be feasible, and it will be the best possible timetable according to the weighted optimization criteria and the allowed computational time. In this way, MOM can also be used as an “any-time scheduler”.

4 Main capabilities of MOM

MOM can optimize existing timetables, schedule and obtain automatically optimized timetables for new periodic and/or non-periodic trains and make them compatible with the existing scheduled trains. The system can also be used to reschedule timetables according to disruptions and delays in on-line, real-traffic management as well as validate and perform capacity analysis.

4.1 Off-line railway planning

MOM obtains optimized railway timetables in accordance with commercial/user requirements, railway infrastructure, traffic constraints, and optimization multi-criteria. MOM is mainly configured as a computer-based support system for railway scheduling. Hence, the user can easily adjust several parameters (traffic constraints, infrastructure data, train data, journeys, etc.) and obtain correct and optimized timetables with different properties, allowing the user to make the final best selection. In addition, a given timetable could be interactively modified by the user, which can use MOM in order to validate its correctness.

4.1.1 Flexible optimization

MOM implements two optimization modes *Standard Mode* and *Flexible Optimization Mode*, which can be selected by the user in each optimization process. *Standard Optimization* (which is the default mode) implies the satisfaction of all the operator, traffic, and infrastructure constraints. In the *Flexible Optimization* mode, some constraints are considered as *Soft Constraints*, so that they can be relaxed if the resulting timetable is considerably improved. For instance, the travel time of a given train between two stations would be diminished by 3% in order to avoid a crossing operation with a train going in the opposite direction. Alternatively, the reception time could be relaxed 30 seconds in that station.

The constraints that can be relaxed in the flexible optimization mode are: the running time of a train in sections, and the temporal margins for reception, expedition, and headway. The user establishes the allowed relaxation bounds for these constraints.

The flexible optimization is applied by MOM if and only if (i) the user selects the *Flexible Optimization* mode, and (ii) the relaxation of the allowed soft constraints is justified with the improvement achieved in the timetable. After a flexible optimization is performed, a set of constraints is proposed to be relaxed. The user can select to validate all of the relaxed constraints, a subset, or none of them, according to viability of the proposed relaxations and the achieved improvement. All the relaxed constraints in a final timetable have to be validated by the user.

Figure 3a, illustrates a feasible timetable obtained by an expert planner, with a total travel time of 46:41:30. This timetable has been automatically optimized by applying the flexible optimization module of MOM. The total travel time is reduced by 5:36:05, (total travel time: 41:05:00) by relaxing some constraints (Figure 3b).



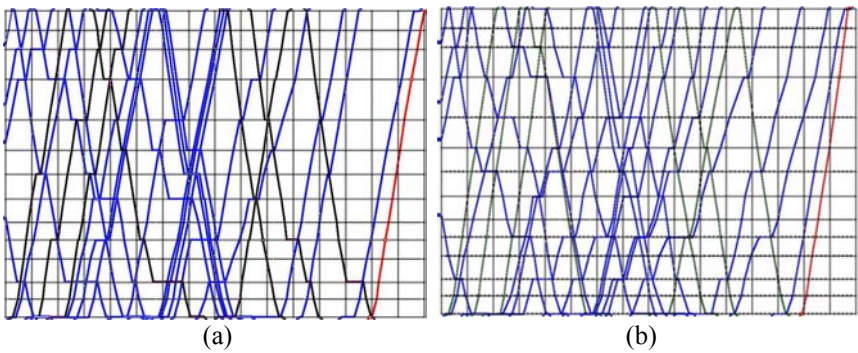


Figure 3: (a) A Standard-Optimized timetable. (b) Improved timetable using *Flexible-Optimization*.

4.2 On-line railway planning

MOM can be used to reschedule traffic timetables according to disruptions and delays in on-line and real-traffic management. The system supports the user in the decision making when disruptions occur. Two objectives can be selected in order to obtain an updated timetable after disruptions: (i) the minimization of the difference between the original timetable and the new timetable for each station and (ii) the minimization of the difference in scheduled arrival time of trains to the final station of their journeys. Figure 4 shows four incidences in a sample instance: two trains that are behind schedule and two trains that are ahead of schedule. After the application of the on-line railway planning module of MOM, the original timetable is updated taking into account the new situation.

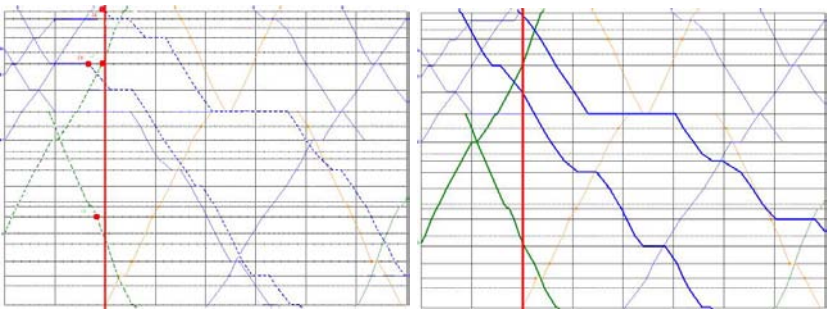


Figure 4: On-line scheduling.

4.3 Railway capacity analysis

Optimizing the use of railway infrastructure is a complex and difficult task. Questions such as “What part of extra traffic can be absorbed by the existing



efficiently evaluate different scheduling parameters and alternative timetables. The maximum optimization level obtains higher quality timetables when the computational time is longer. Figure 6 shows the optimization level with respect to the computational running time of three similar examples to the previous one.

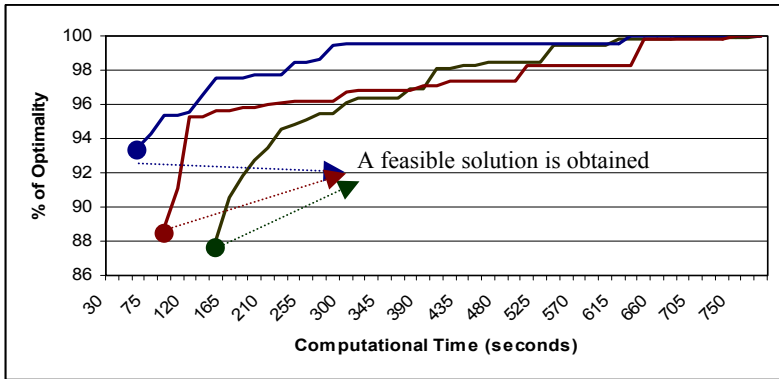


Figure 6: Optimization level vs computational time.

6 Conclusions

A Railway DSS is presented in this work. The system has three main scheduling modules:

- i) Off-line Railway Planning, where the system solves and obtains optimized railway timetables in accordance with: commercial/user requirements, railway infrastructure, traffic constraints, and optimization criteria,
- ii) On-line Railway Planning, which reschedules traffic timetables when disruptions, incidences, or delays occur, and
- iii) Railway Capacity Analysis, which validates and performs capacity analyses of railway infrastructure.

MOM can be integrated in existing computer-based tools and obtains the required data from standard databases and friendly interfaces. This system is currently being used successfully by the Spanish Manager of Railway Infrastructure (ADIF) for obtaining off-line timetables. Currently, MOM is being improved adding the consideration of the analysis of the robustness and the stability of timetables generated. Further features of MOM can be obtained from <http://www.dsic.upv.es/users/ia/gps/MOM>.

Acknowledgements

This work has been partially supported by the research projects TIN2004-06354-C02- 01 (Min. de Educación y Ciencia, Spain-FEDER), FOM-70022/T05 (Min.



de Fomento, Spain) and FP6-021235-2, IST-STREP (UE). We appreciate the assistance of Jose Estrada (Spanish Administration of Railway Infrastructure - ADIF).

References

- [1] Barber, F., M. A.Salido; L. Ingolotti; M. Abril; A. Lova; P. Tormos., *An Interactive Train Scheduling Tool for Solving and Plotting Running Maps*. Current Topics in Artificial Intelligence. Lecture Notes in Computer Science/LNAI Series. Springer-Verlag. Vol. 3040, pp. 646-655, 2004.
- [2] Hürlimann, D., "Object oriented modelling of infrastructure elements and business processes in railways". Ph D. Thesis. ETH Zurich.Switzerland, 2001.
- [3] Ingolotti,L., F.Barber, P. Tormos, A. Lova , M. A. Salido, M. Abril, *An Efficient Method to Schedule New Trains on a Heavily Loaded Railway Network*. Advances in Artificial Intelligence. Lecture Notes in Computer Science/LNAI Series. Springer-Verlag Vol. 3315. pp. 164-173, 2004.
- [4] Middelkoop, D. Bouwman, M. "Train network simulator for support of network wide planning of infrastructure and timetables. In: Computer in Railways VII, WIT Press, pp. 267-276. 2000.
- [5] Salido, M.A., M. Abril , F. Barber, L. Ingolotti, P. Tormos, A. Lova., *Topological Constraints in Periodic Train Scheduling*. Frontiers in Artificial Intelligence and Applications. Planning, Scheduling and Constraint Satisfaction: From Theory to Practice. Frontiers in Artificial Intelligence and Applications. IOS Press. Vol. 117, pp. 11-20, 2005.



Measurement of train driver's brain activity by functional near-infrared spectroscopy (fNIRS)

T. Kojima¹, H. Tsunashima² & T. Y. Shiozawa³

¹*Graduate School of Nihon University, Japan*

²*College of Industrial Technology, Nihon University, Japan*

³*School of Medicine, Nihon University, Japan*

Abstract

Train drivers are strongly urged to avoid making human errors of judgment. To avoid human error in train operation, driving support systems should be developed in consideration of human behavior. In developing such systems it is important to understand the relation between the train operation and the brain activity of driver. The brain activity during the driving operation with a train-driving simulator is measured by a functional near-infrared spectroscopy (fNIRS). To compare the changes in the brain activity with the driving operation, we developed an analysis method that can extract the brain activity related to task by using a wavelet-based multi-resolution analysis. The results show that the brain function due to a train driving can be evaluated with the proposed method.

Keywords: railway, human factors, brain function, near-infrared spectroscopy, wavelet, multi-resolution analysis, train simulator.

1 Introduction

One of the characteristics of train driving is that the driving operation tends to be more monotonous than car driving. This decrease in the driver's arousal level may cause human error. Railway companies provide education and training to train crews as part of measures to prevent such human-error-related accidents. In addition, companies have introduced backup systems such as Automatic Train Stop (ATS) devices. On some railways, Automatic Train Operation (ATO) devices have been introduced to automate train operation.



However, this increased automation requires train drivers to focus on monitoring tasks and may thus cause a worrying decrease in driver arousal levels. In the automotive field, drowsy driver detection and warning systems, which can detect drowsy driving on the basis of information about the driver's physiological state and/or driving behavior and give a warning, have been studied. Basic studies on human behavior have been conducted in the field of train operation, but these studies have not yet actively contributed to train operation support. To avoid human error in train operation, we consider that the development of driving support systems should take human behavioral characteristics into account.

In developing such systems, it is important to understand: 1) how train drivers recognize information, 2) what judgments they make on the basis of this information, and 3) what actions they take. A driving simulator that can measure the driver's physiological state or the like under various conditions is needed to analyze the driver's actions.

This paper describes the measurement of brain function during train operation using a functional near-infrared spectroscopy (fNIRS), and proposes a new analysis method for extracting the brain activity related to a driving operation by wavelet-based multi-resolution analysis.

2 Development of train-driving simulator for evaluation of human factors

For human factor studies, a train simulator is needed which can simulate realistic train operation. We developed a train-driving simulator which can change the environment flexibly to suit the purpose of the experiment.

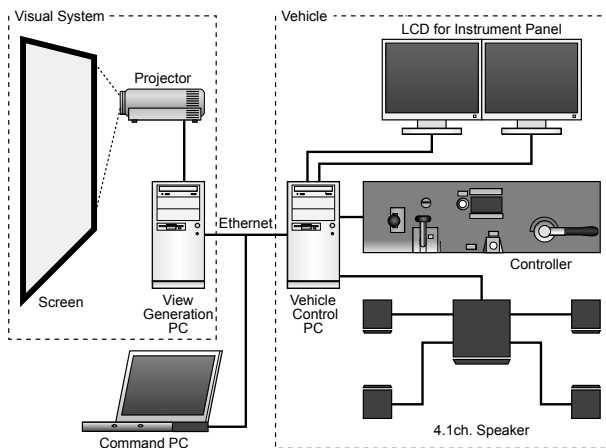


Figure 1: System configuration of train-driving simulator.

Fig. 1 shows the system configuration of the driving simulator that was developed. The simulator consists of three units: 1) a vehicle unit, composed of a

vehicle control computer, a train cab controller, monitors for instrument panel, and speakers; 2) a visual system unit, composed of a view generation computer, a projector, and a screen; and 3) a command computer. Once the handle positions were input to the train control computer, actions such as the calculation of train movement, the generation of sounds, and the creation of instrument panel images on the monitor were carried out on the basis of the train's properties. When the calculated train position data were sent to the view generation computer, computer graphic view images were created and the images were projected by the projector onto the screen located at the front of the cab. The command computer could perform a number of actions: for example, it could cause accidents to occur and change stations.

3 Measurement of brain function

3.1 Recent trends in brain function examination and the need for measurement during driving operations

If we are to understand the mechanism of a potential accident, we need to study the brain activity of the person who might cause the accident, as representative of the person's physiological condition. Functional magnetic resonance imaging (fMRI) is attracting attention as a noninvasive method of diagnostic imaging. fMRI examination, which uses magnetism to enable the observation of localized changes in oxygenated hemoglobin levels in the brain, has contributed greatly to the clarification of brain functions such as language and cognition. However, fMRI has various limitations and is difficult to use in the evaluation of driving operations. For example, the examinee has to lie supine in a narrow cylinder, without moving his/her head during the examination.

In contrast, near-infrared spectroscopy (NIRS) has recently become popular. This is also noninvasive examination method. A relatively small type of NIRS with a single channel is has been applied to measurements during airplane operation [5]. This single channel NIRS can be used to observe the circulatory dynamics of only a small area in the brain. Functional near-infrared spectroscopy (fNIRS) [3, 4, 6], so-called optical topography, allows the mapping of wide area and evaluation of functional localization in the cerebrum using NIRS technology.

Either type of NIRS apparatus is characterized by the stable capture of circulatory dynamics in the brain during body movement in real time and promises to be useful for such evaluations.

3.2 Principle of near-infrared spectroscopic measurement

Fig. 2 shows a diagram of near-infrared spectroscopic measurement of cerebral blood flow. Near-infrared rays enter human tissues through optical fibers, propagate in the tissues and radiate from the tissues. NIRS detects the radiated rays, and measures relative variations of oxygenated hemoglobin (oxy-Hb) and deoxygenated hemoglobin (deoxy-Hb) based on those absorbencies.



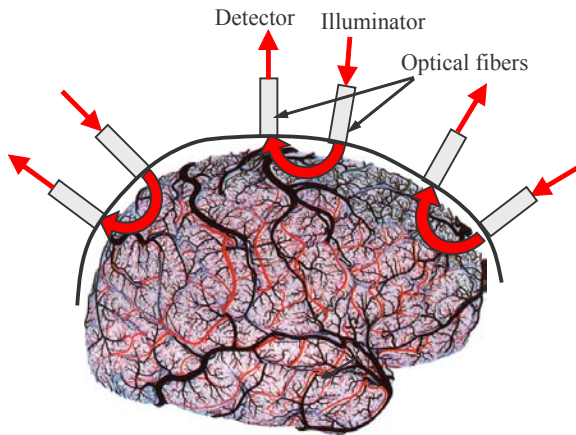


Figure 2: Diagram of near-infrared spectroscopic measurement.

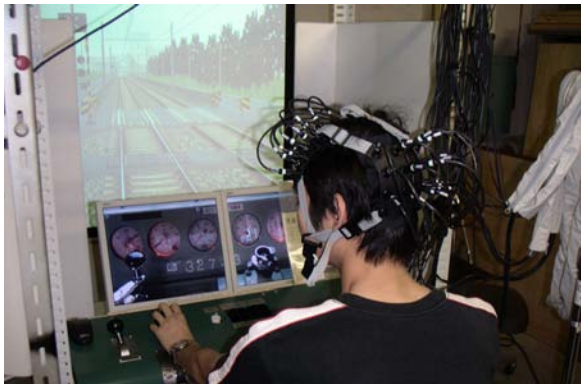


Figure 3: Experiment of measurement of brain function using train-driving simulator.

3.3 Measurement of brain function using the train-driving simulator

3.3.1 Experimental methods

We measured brain activity using the train-driving simulator. Fig. 3 shows how the experiment was carried out. We simulated a length of 2.3 km with three stations on a real railway line. A station-stop as long as 50 s was set in order to measure changes in cerebral blood flow during a stop. Three runs were made, with one run set at 6 min. The examinees were two males who had not driven real trains but were fully proficient in operating the simulator. The examinees were instructed to drive according to the signals and indicators, following a timetable. The measurement instrument used was a near-infrared imaging system (OMM-3000, Shimadzu Corporation, Japan) as shown in fig. 4. Fig. 5 shows the arrangement of optical fibers and measurement positions. Numbers represent measurement channels. Twenty-two (22) channels were set in the frontal region and 22 in the occipital region, giving total of 44 channels.





Figure 4: A near-infrared imaging system.

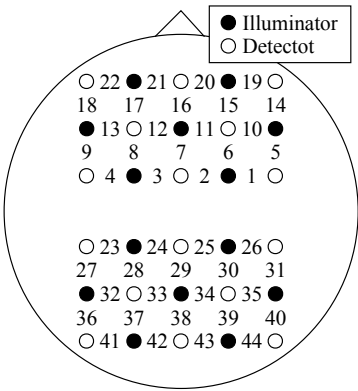


Figure 5: Arrangement of optical fibers and measurement positions.

3.3.2 Measurement results

Fig. 6 shows the results from all the channels in the frontal region, measured just after starting while the driver was in the section with 3 stations. The general trend was that the oxy-Hb concentration increased and the deoxy-Hb concentration decreased over the course of the operating time, indicating activation of the brain. Similar activation was observed in the occipital region. Furthermore, changes in hemoglobin concentration were observed in response to tasks of stopping at stations. The result of channel 7 which shows remarkable task-related changes is shown in Fig. 8. However, fNIRS signals include measurement noise and any changes in brain activities which are not related to the driving operation. To compare the changes in brain activity with individual driving operations, it is necessary to extract the task-related signal. We propose an analysis method using a wavelet-based multi-resolution analysis.

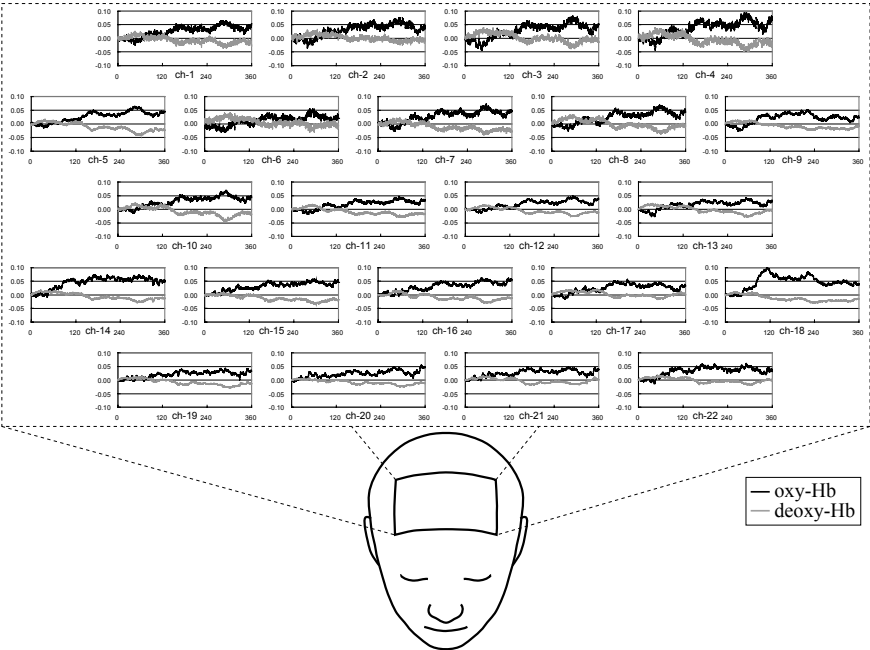


Figure 6: Changes in hemoglobin concentration in frontal.

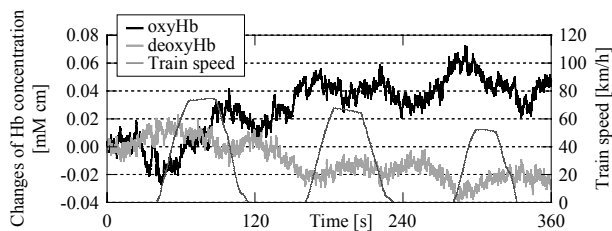


Figure 7: Changes in hemoglobin concentration on channel 7 and train speed.

4 Decomposition and reconstruction of fNIRS signal by wavelet-based multi-resolution analysis

4.1 Wavelet transform

In wavelet transform [2], a signal $S(t)$ is described by shifting and scaling a small wave, $\psi(t)$, called a mother wavelet. The signal is analyzed on the basis of the mother wavelet transform. A continuous wavelet transform (CWT) is given by:

$$(W_{\psi}S)(a,b)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{a}}\overline{\psi\left(\frac{t-b}{a}\right)}S(t)dt \tag{1}$$



$\psi((t-a)/b)$ shifts the time (phase) by b and makes the frequency $1/a$. This continuous wavelet transform, however, requires a considerable amount of computation, and information is redundant. Therefore, by making $a=2^m$ and $b=2^m n$ in eqn (1), we obtain:

$$D_{m,n} = \int_{-\infty}^{\infty} S(t) \psi_{m,n}(t) dt \quad (2)$$

where:

$$\psi_{m,n}(t) = 2^{-m/2} \psi(2^{-m}t - n) \quad (3)$$

which is called a discrete wavelet transform (DWT). This discrete wavelet transform is characterized by more efficient translation of signals.

4.2 Multi-resolution analysis

Multi-resolution analysis (MRA) decomposes a signal into a number of components at different resolution by using the discrete wavelet transform. A time history $S(t)$ is decomposed into some detailed (high-frequency) components and an approximated (low-frequency) component. The original signal $S(t)$ given by:

$$S(t) = S_{m0}(t) + \sum_{m=-\infty}^{m0} d_m(t) \quad (4)$$

where d_m are detailed components:

$$d_m(t) = \sum_{n=-\infty}^{\infty} D_{m,n} \psi_{m,n}(t) \quad (5)$$

In the wavelet transform, the choice of a mother wavelet $\psi(t)$ is important. We employed *Daubechies* wavelet [1], which is orthonormal base and compactly supported wavelet. The vanishing moments of *Daubechies* wavelet can be changed by index N . We decided to use a relatively high-order generating index, $N = 7$ (fig. 8).

4.3 Decomposition and reconstruction of fNIRS signals

Fig. 9 shows the result in multi-resolution analysis of the decomposition level of ten (10) for oxy-Hb signal on channel 7 (shown in Fig. 7). Fig. 8 indicates that the trend of all the operation subjects was extracted in the approximation component (a_{10}), and that other changes were decomposed into the detail components. Components d_1-d_3 had comparatively large amplitudes, which we considered were caused by heartbeat-related blood flow and measurement noise. Large changes were observed in components d_9 and d_{10} . These were considered to be task-related changes, because the repetition interval of the starting and stopping at stations is 120 s, which is in the frequency band of component d_{10} . We added components d_8 , d_9 and d_{10} to reconstruct.



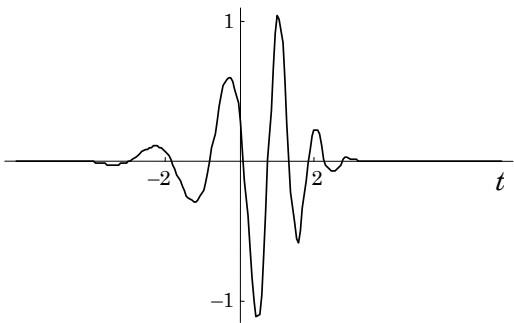


Figure 8: *Daubechies* wavelet ($N=7$).

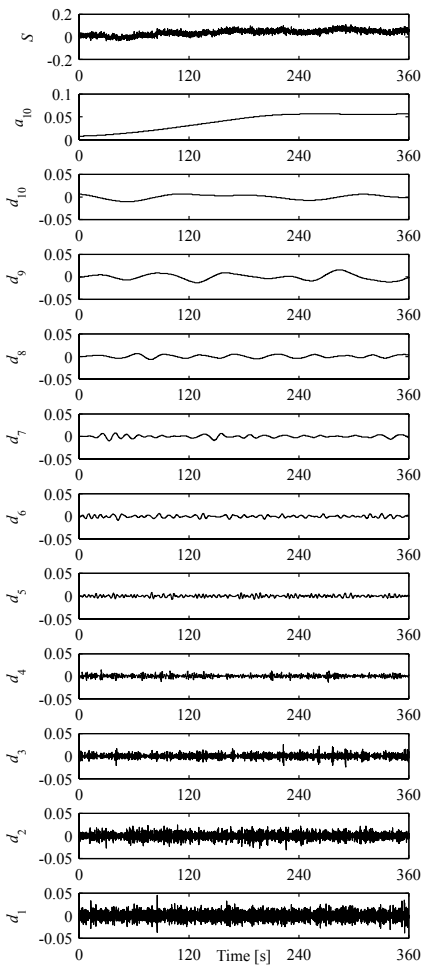


Figure 9: Multi-resolution analysis of oxy-Hb signal on channel 7.



Fig. 10 shows the comparison of the reconstructed signals on channel 7 (shown in Fig. 7) and train speed. oxy-Hb concentration was highest when departing from stations, and deoxy-Hb showed the opposite trend. This confirmed activation of the brain during stops.

After the reconstruction for every measured channel, we created the functional brain images in frontal. The results are shown in fig. 11. We consider this imaging is very effective in evaluating driving behavior.

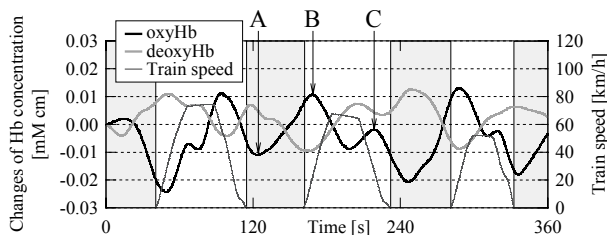


Figure 10: Reconstructed signals of changes in hemoglobin concentration on channel 7.

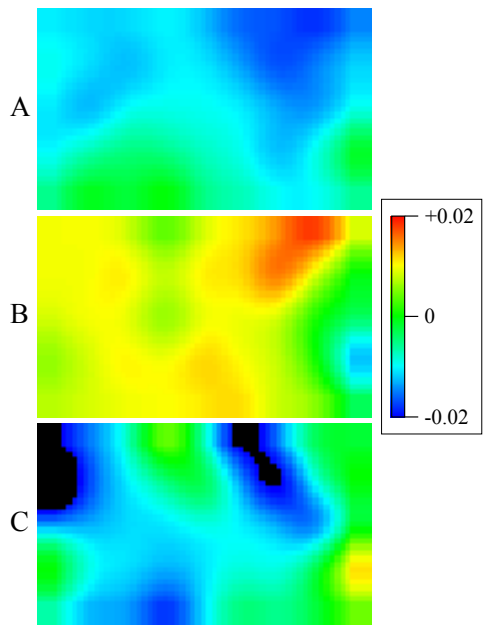


Figure 11: Functional brain imaging in frontal.

5 Conclusions

We measured brain function by using functional near-infrared spectroscopy (fNIRS) during simulator operation. The results show that oxy-Hb concentration



increased and deoxy-Hb decreased with driving, and that train driving activated the brain. Additionally, multi-resolution analysis using wavelet transform is useful for evaluating brain activity in response to individual driving operations.

We are planning a future investigation into the physiological evaluation of changes in cerebral blood flow by making simultaneous measurements and comparisons of fMRI, brain waves and heartbeat.

References

- [1] Daubechies, I., Orthonormal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics*, **41(7)**, pp. 909–996, 1988
- [2] Daubechies, I., Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series In Applied Mathematics: *Society for Industrial and Applied Mathematics*, **(61)**, 1992
- [3] Eda, H. et al, Multi-channel time-resolved optical tomographic imaging system, *Rev Sci Instrum*, **(70)**, pp. 3595–3602, 1999
- [4] Jöebssis, FF., Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters, *Science*, **(198)**, pp. 1264–1267, 1977
- [5] Kobayashi, A. & Miyamoto, Y., In-flight cerebral oxygen status: continuous monitoring by near-infrared spectroscopy, *Aviat Space Environment Med*, **(71)**, pp. 177–183, 2000
- [6] Villringer, A. & Chance, B., Non-invasive optical spectroscopy and imaging of human brain function, *Trends Neurosci*, **(20)**, pp. 435–442, 1997



Methodology for the LCC-Analysis and the optimal migration of the railway operations control on the example of ETCS

M. Obrenovic, B. Jaeger & K. Lemmer

*German Aerospace Center, Institute of Transportation Systems,
Braunschweig, Germany*

Abstract

In order to reduce the number of approximately 20 different train control systems in Europe, and thus to realise cost effective and seamless cross-border railway operation, The European Train Control System (ETCS) will be implemented within the next years. Besides the given legislative preconditions, the cost efficiency as well as the optimal migration strategy, both on an operative and strategic level, will be crucial for the commercial success of the system.

The Life-Cycle Cost Analysis (LCC) takes into account the overall costs appearing during the entire life cycle, considering all purchasing, owning and the disposal of the system. Additionally it identifies cost drivers by analysing the “cause-and-effect chain” between relevant figures and processes. Therefore, a system breakdown structure plan and a cost breakdown structure plan are mapped on the different phases in the life-cycle of the analysed operations control system. Furthermore, a process analysis needs to be carried out in order to identify and to optimise those areas where cost-drivers are generated, e.g. the development or the maintenance process. Based on the LCC Analysis and additionally on a Multi-Criteria Analysis, considering operational, technical and other aspects, the optimal target state will be identified.

In a last step, migration scenarios for the selected system need to be developed and evaluated systematically. Thus, the optimal migration strategy can be carried out. Furthermore, the migration costs and the duration of the process have to be considered beforehand in the LCC Analysis. As a result, a closed loop and an interaction related to both thematic areas – LCC and the migration of the railway operations control – are found.

The Institute of Transportation Systems has developed a consistent methodological framework including tool-support, which is covering this process and will be demonstrated on ETCS in this contribution.

Keywords: migration, Life Cycle Costing, ETCS.



1 Introduction

The migration of innovative technologies into the complex railway system is characterised by a number of heterogeneous criteria and constraints. Due to the shared functionalities and components allocated on the track-side as well as the rolling stock, particularly the railway operations control raises high requirements on the process. Mayor goals of the ETCS migration are e.g. a certain level of operational and RAMS (Reliability, Availability, Maintainability, Safety) performance as well as low LCC and migration costs whilst providing technical and operational interoperability. Additionally, train circulation on European corridors and the mix between passenger and freight traffic are to be considered. Thereby, a multi-dimensional optimisation problem has to be solved. This requires a tool-supported methodology dealing with the migration in the domain of railway systems. [2] Figure 1 shows the process model based on the migration drive to change the current system, the selection of the appropriate target state as well as the development and evaluation of migration strategies.

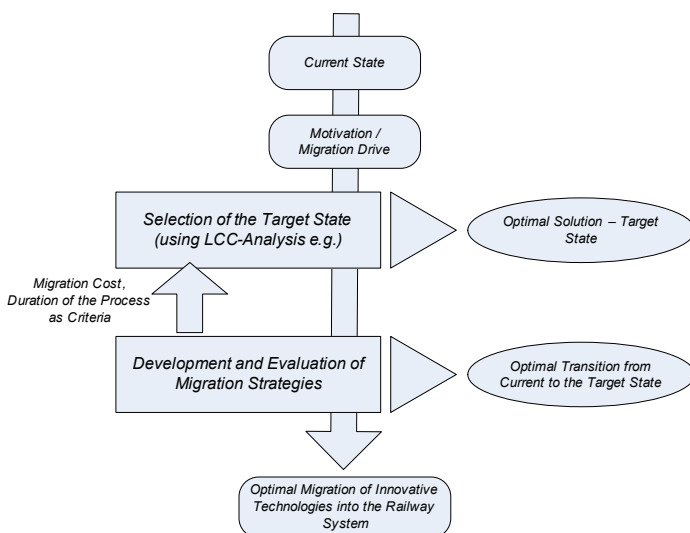


Figure 1: Interaction in the process model for the migration of new technologies.

Expecting the LCC to be one of crucial criteria for system selection, and bearing in mind that migration cost and duration are to be considered with the LCC-Analysis already, the closed loop and the interaction between the selection of the target state and the migration in the overall process model become evident.



2 European Train Control System – a new technology for European railways

Currently approximately 20 different train control systems are in use in Europe. In order to reduce this number, and thus to realise cost effective and seamless cross-border railway operation, ETCS will be implemented within the next years. Therefore, the legislative preconditions in the form of EU directives 96/48/EG and 2001/16/EG have been established. Besides the political demand, the economical efficiency and the optimal migration of this new technology will be crucial conditions for its success. The migration on the strategic level is described by corridors for the high-speed and conventional rail. Now each country has to find the optimal way for the migration of ETCS on particular corridors, concerning the characteristics of the national railway network as well as the rest of the related corridor and the entire future ETCS network.

Deploying ETCS on the European railway network, among others, the following impacts are expected:

- Decrease of transport time, especially in the cross-boarder traffic;
- Increase of track capacities;
- Increase and harmonisation of the safety level of European railways;
- Lower LCC especially due to
 - Lower asset cost due to the changing of the purchasing market, standardisation and harmonisation;
 - Less equipment, and thus lower maintenance cost track-side;
- Interoperability;
- Opening of the sales markets for the suppliers;
- Barrier-free access to the railway infrastructure.

But on the other hand, high invest for the ETCS migration will be necessary. The problem of national legacy systems and their long life-cycle and thus long term depreciation has to be solved. Financing of the ETCS equipment and certification partially by subsidies for the Infrastructure Managers (IM) and Train Operator Companies (TOC) is necessary. All these questions, as well as the identification of the suitable ETCS Level for certain track sections, have to be answered in order to ensure the commercial success of the system.

3 LCC-analysis as a methodology for the system selection

For the identification of the criteria suitable for the selection of the target system, the specific migration drive has to be regarded, i.e. the motivation to introduce a new technological system into an existing technical environment. There are four basic types of motivation, which can be identified for the migration of railway operations control [2]:

- Rationalisation;
- Advancement of the track performance (e.g. interoperability, speed);
- Absence of the availability of technical components or systems;
- Political decisions.



Due to the high complexity of the decision model, including figures and indicators from the economical, technical, operational or political point of view, the Value Benefit Analysis (VBA) is a suitable method to be applied. It is an approach for the systematic operations research, used at problems of the selection by different project options. The VBA originates from the engineering sciences and allows, contrary to the Cost-Benefit Analysis, the view on non-monetary assessment criteria as well, and aims not only at the economic efficiency. Nevertheless, the economical aspects and especially the LCC of the regarded systems will be crucial criteria in the decision process, receiving the appropriate weight within the scope of the VBA.

The LCC-Analysis takes into account the overall costs appearing during the entire life cycle, considering all purchasing, owning and the disposal of the system. It additionally identifies cost drivers by analysing the “cause-and-effect chain” between relevant figures and processes. Therefore, a LCC-Model is to be elaborated for the regarded operations control system. Furthermore, there is a process analysis to be done in order to identify and optimise the areas where the cost-drivers are generated, e.g. the development or the maintenance process. The Standard DIN EN 60300-3-3 [3] gives a guideline including the elaboration of the system breakdown structure plan as well as the cost breakdown structure plan, needed for the analysis.

LCC-Model is a simplified image of the reality, which abstracts the main characteristics and attributes of the regarded system transferring them into the cost factors. Especially due to the long life-cycle of the railway operations control, these costs have to be estimated partially. The long life-cycle requires the consideration of the product modification costs during the operation phase, too. In the next sections, both the system breakdown structure plan as well as the cost breakdown structure plan for ETCS will be elaborated.

3.1 System breakdown structure plan for ETCS

In order to develop the LCC-Model, the system analysis has to be carried out first. Thereby, the system breakdown structure shows the system components, i.e. units the LCC can be allocated to. Thus, the first step for the identification of cost drivers is done. Separated in three parts

- Track-side components;
- GSM-R;
- On-board components;

Fig. 2 shows the system structure of ETCS. Thereby, we renounce a more detailed view – this level seems sufficient in order to present the methodology.

LEU	Lineside Electronic Unit
RBC	Radio Block Center
GSM-R	Global System for Mobile Communication – Railway
OBU	On-board Unit



EVC	European Vital Computer
JRU	Juridical Recording Unit
TIU	Train Interface Unit
BTM	Balise Transmission Module
RTM	Radio Transmission Module
LTM	Loop Transmission Module
STM	Specific Transmission Module / Translation of the National System
MMI	Man-Machine Interface / Driver-Machine Interface

Once the system is analysed, the possible cost positions, arising during the life-cycle of the regarded CC-System (Control-Command), are to be identified. Therefore, in the following section a cost structure will be elaborated. Thus, single cost positions occurring during the life-time can be mapped on the system components.

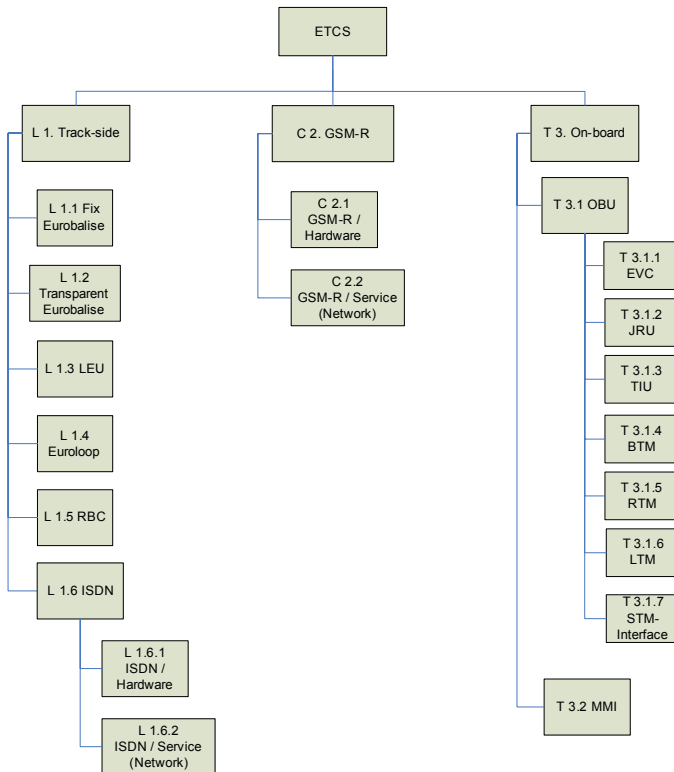


Figure 2: System breakdown structure of ETCS.

3.2 Cost breakdown structure plan

From the user point of view – Infrastructure Managers and Train Operator Companies – three basic phases, specifying the LCC, can be identified (Fig. 3).



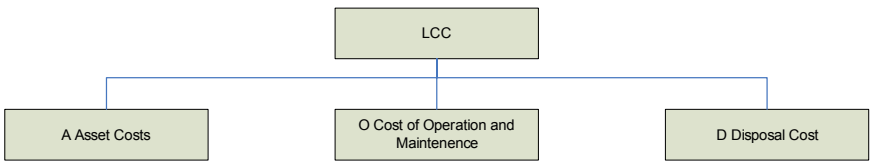


Figure 3: Layer of the LCC breakdown structure – operator’s point of view.

In this contribution, the design and the development of the system won’t be regarded. For system users, these costs will be expressed in the purchase price and the usage of the system during the operation. Nevertheless, these phases are crucial for the determination of the LCC. Thus, it is highly recommended to optimise the development process, bearing in mind that approx. $\frac{3}{4}$ of the LCC are being determined there.

In the next sections, costs appearing during the three phases of the user-oriented life-cycle will be presented.

3.2.1 Asset cost – A

The classical investment appraisal uses asset costs as a key criterion for the system selection. In the LCC-Analysis, it is only one part of the chain. Fig. 4 shows the elements of the asset costs.

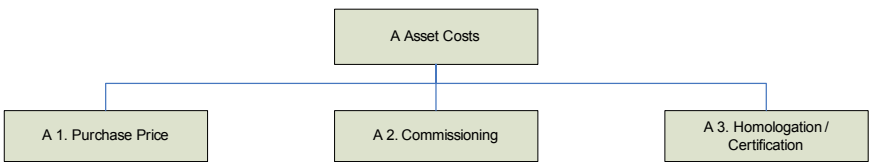


Figure 4: Layer of the asset cost breakdown structure.

Table 1: Asset costs.

A 1	Purchase Price
A 2.1	Opportunity Costs of Commissioning – Downtime, etc.
A 2.2.1	Personnel Training Costs for Commissioning
A 2.2.2	Personnel Installation Costs of Commissioning
A 2.3.1	Material Costs of Commissioning - Supplies
A 2.3.2	Material Costs of Commissioning - Facilities
A 2.4	Planning / Programming Costs of Commissioning
A 3	Homologation / Certification

3.2.2 Cost of operation and maintenance – O

Especially during the long-term operation phase, various costs can be identified (Fig. 5, Table 2).



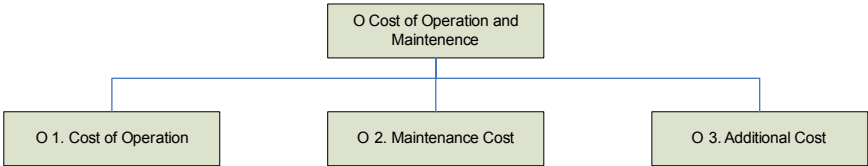


Figure 5: Layer of the cost of operation and maintenance breakdown structure.

Table 2: Cost of operation and maintenance.

O 1.1.1	Material Cost of Operation - Supplies
O 1.1.2	Material Cost of Operation – Facilities
O 1.2	Personnel Cost of Operation
O 1.3	Occupancy Cost of Operation
O 2.1.1	Personnel Cost of Preventive Maintenance
O 2.1.2	Costs for Spare Parts of Preventive Maintenance
O 2.1.3	Tool Costs of Preventive Maintenance
O 2.2.1	Personnel Cost of Repair
O 2.2.2	Costs for Spare Parts of Repair
O 2.2.3	Tool Costs of Repair
O 2.3	Opportunity Cost for Downtime
O 3.1	Cost of Capital
O 3.2	Cost of System Modification

Due to the strong impact on the entire LCC, in this phase the processes determining the cost indicators and figures are to be analysed in-depth. Thereby, the impact of key figures like Mean Time Between Maintenance (MTBM), Mean Time To Repair (MTTR) or Mean Time To Failure (MTTF) has to be regarded. Modelling of the influence of the railway operation or the internal processes on these figures allows the identification of possible cost drivers and thus the identification of optimisation areas.

3.2.3 Disposal cost – D

In spite of the extremely long life-time of railway operations control, the costs related to the recycling are to be considered, too (Fig. 6).

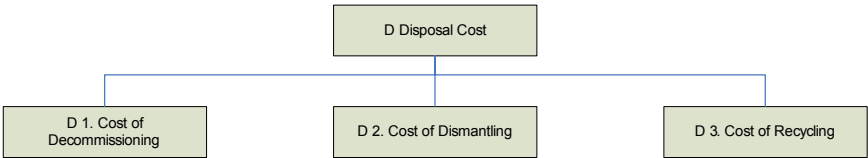


Figure 6: Layer of the disposal cost breakdown structure.



Table 3: Disposal cost.

D 1.1	Personnel Cost of Decommissioning
D 2.1	Personnel Cost of Dismantling
D 2.2.1	Material Cost of Dismantling – Supplies
D 2.2.2	Material Cost of Dismantling – Facilities
D 3	Cost of Recycling

3.2.4 Overall LCC

In order to determine the overall LCC, the single cost positions presented in the sections above have to be accumulated.

$$LCC = A_0 + \sum_{j=1}^n (O_j \times \frac{1}{(1+d)^j}) + D_n \quad (1)$$

LCC	Life Cycle Costs (User-Oriented)
A_0	Positions of Asset Cost to the Date 0
O_j	Positions of Cost of Operation and Maintenance during the Life-Cycle
D_n	Positions of Disposal Cost to the Date n
$j \dots n$	Life-Time of the System
d	Discount Rate

To consider the net-present-value, the costs are to be discounted to a defined rate. Using this approach, an efficient method for the assessment of economical aspects in the selection process can be applied.

4 Development and assessment of migration scenarios

Once the appropriate system is selected, an optimal strategy for the migration from the current state into the target state has to be elaborated. Here, the transition from various national CC-Systems toward ETCS is meant.

Especially due to the safety relevant interaction of the track-side and the on-board system components of the railway operations control, the development of migration scenarios is a complex process. The crucial requirement is to keep up operation during the migration period. Considering the fact that ETCS is not compatible to the national CC-Systems, which have to be replaced, parallel equipment for a certain period is necessary.

The final target of the ETCS migration is not to have one more train control system overlaid to the national systems. Rather, ETCS has to be established as a substitution of national CC-Systems, at least on defined corridors. This should be a requirement on the migration process, while having in mind the higher operational and maintenance costs (as parts of LCC) in case of the parallel equipment track- or train-side.

ETCS migration raises problems on different levels. On the one hand, we have a strategic question of ETCS deployment using the corridor strategy. On the



other hand, the characteristics of certain tracks and sections have to be considered on the operative level. Assessment of different migration strategies, e.g.:

- Parallel equipment with ETCS and the national CC-System track-side;
- Parallel equipment on-board;
- Parallel equipment on both system sides;
- STM-Strategy;

includes the analysis and optimisation of two crucial figures: *cost and time*.

Based on the published data on the one hand and some assumptions being necessary due to the low maturity of the system on the other hand, Fig. 7 shows one of the possible ETCS Level 1 strategies. It is the parallel equipment train-side, including the system deployment (infrastructure and rolling stock) and the annual cost trend. Thus, the identification of the “break-even-date”, realising annual (not general) benefits of the ETCS migration, can be determined. See detailed calculation in [1].

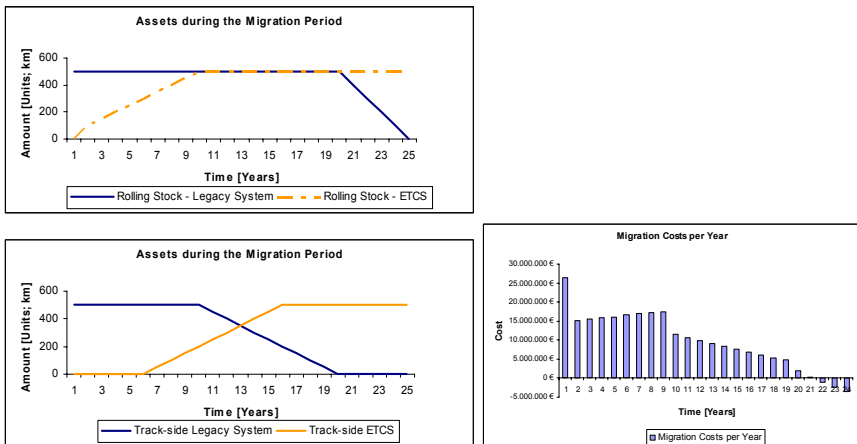


Figure 7: Example: parallel equipment train-side – assets and annual cost trend [1].

Migration cost in each scenario is calculated by comparing the asset costs and the cost of operation and maintenance of ETCS with the cost of the reference scenario, keeping the legacy system [4].

As a part of the methodology, at the Institute of Transportation System of the German Aerospace Center, a tool-support for the evaluation of migration scenarios is provided as well.

Migration cost has to be considered already in the LCC-Analysis, carried out in the selection phase. Due to the recursive dependence of both phases – selection and migration – there is a closed loop regarding the general methodology. Normally, having a small number of alternatives in the domain of railway operations control and especially CC-Systems, the solution space and thus the global problem is manageable.



5 Conclusions

Economical efficiency and the optimal migration of innovative technologies in the railway domain are key conditions for their success. Currently, in spite of the political pressure, the pace of deployment of the new European Train Control System will significantly depend on these aspects. Therefore, a methodical determination of impacts on the one hand and the optimisation on the other will be necessary to convince the IMs and TOCs of the system. To make the decision on the implementation at all, respectively on the appropriate ETCS level on certain track sections, LCC-Appraisal should be applied within the scope of the Value-Benefit-Analysis. Thereby, the elaboration of the system breakdown structure and the cost breakdown structure is the basis for the LCC-Model. From the users' point of view, costs of purchasing, owning as well as the disposal of the system have to be considered. Thus, single cost positions during the life-time and the cost drivers can be identified. Based on this information on the one hand and by comparing the alternatives, a selection of the optimal system solution can be made. On the other hand, the optimisation areas become evident.

Solving the optimisation problem, occurring with respect to the migration of railway operations control, deployment of ETCS will be accelerated. To minimise additional cost and operational constraints, a methodical approach has to be applied in the process of the development and evaluation of migration strategies. Furthermore, both the strategic as well as the operative level of the migration on the one side and the claims of IMs as well as TOCs on the other side have to be considered.

From the economical point of view, there is a closed loop between the LCC-Analysis and the optimal migration. Integration of both aspects in a consistent methodology, as shown in this contribution, ensures a suitable approach, covering the questions of economical efficiency and the migration of railway operations control.

References

- [1] Obrenovic, M., Jaeger, B., Lemmer, K., *Development and Evaluation of Migration Strategies for ETCS Level 1 Applied on a Synthetic Track*, ZEL 2006, Zilina (SK) – accepted
- [2] Knollmann, V., Obrenovic, M., Jaeger, B., Lemmer, K., *Optimized Migration Process in Rail Traffic Using Simulation Support*, 1st International Seminar on Railway Operations Modelling and Analysis, RailDelft2005, Delft (NL)
- [3] DIN EN 60300-3-3 Dependability management – Part 3-3: Application guide – Life cycle costing (IEC 60300-3-3:2004); German version EN 60300-3-3:2004
- [4] Obrenovic, M., Meyer zu Hoerste, M., Jaeger, B., *Phase Model for the Development and Evaluation of Migration Strategies in Rail Traffic*, The 5th European Congress and Exhibition on Intelligent Transport Systems and Services (HITS), Hannover 2005



Application of location detection system using active type RFID tags to railways

K. Seki, S. Suzuki, M. Ukai & R. Tsuchiya
Railway Technical Research Institute, Japan

Abstract

As active type RFID tags can transmit their identification information about 10 m by radio, we can calculate locations of persons who hold tags with an accuracy of a few meters, by installing tag receivers appropriately and by the function of a location calculating server which analyzes the gathered information regarding receivers which have received the tags' identification information. We are developing two application systems to make use of active type RFID tags. The first one is an information provision system for passengers. In this system, passengers hold tags, location calculating servers are installed at stations and in trains, and an information delivering server is located at a management center. The location calculating servers at stations always transmit calculated location information of tags to information delivering server via an Intranet. On the other hand, the location calculating servers in trains store the calculated location information while trains are running between stations and transmit them at a stroke via communications lines such as wireless Local Area Networks to the information delivering server when the trains stop at stations. Therefore, the information delivering server can trace the locations of passengers continuously and manage their status such as entering into stations and getting off trains almost on time. Accordingly, it can deliver guidance information to mobile phones of the passengers via e-mail at appropriate places and timing. The second one is a system to cope with abnormal situations. In this system, station staffs hold tags. When a video surveillance system detects an illegal event such as violence and suspicious loitering, it transmits an alarm to the nearest staff to directly go to the location. As the locations of the staff are detected by tags indoors and by Global Positioning System (GPS) outdoors, they can be continuously traced wherever they are.

Keywords: active type RFID tag, Information provision system, video surveillance system.



1 Introduction

In the future ubiquitous society, grasping locations of persons is one of the most essential things in order to understand their located situations and provide the most suitable services to them. Locations information of railway workers is also important to improve efficiency and rapidness of daily works. Grasping their positions at any places and tracing their movement continuously, however, have not been realized yet, because in a railway system there are so many varieties of environments where locations of persons are to be detected such as indoors, outdoors, and on-board. Though a lot of technologies to detect locations are proposed and realized, it is impossible for only one method to cover all places and requirements.

In this paper, we focus on active type RFID tags and propose two application systems. First we survey a few technologies to detect locations of persons and compare their characteristics. We then explain the first application system, which detects locations of passengers and provides appropriate information to them according to their locations. We tentatively produced this system and verified that the system worked well. Next, we explain the second application system, which grasps locations of station staffs by both active type RFID tags and GPS, detects some abnormal events, and alarms the staffs nearest to the detected event. Finally, we summarize the discussion as the conclusion.

2 Location detection of persons in railways

In this chapter, we survey examples of technologies to detect locations of persons in railways. As these technologies have their own merits and demerits, it is thought to be effective to develop an appropriate system by combining these technologies.

2.1 Global Positioning System (GPS)

GPS is available outdoors with the precisions of about 15 m, and commercial services to guide persons who hold cellular phones with GPS receivers are emerging. However, GPS is difficult to use indoors and on-board, and the precision may be worse or even the location detection itself is impossible when positions of satellites are inappropriate.

2.2 Entering and exiting data of Automatic Fare Collection (AFC) gates

By using entering and exiting data that AFC gates record, it is possible to grasp entering or exiting a station of a passenger without any action by a passenger except inserting or touching a ticket to a gate. Passenger guidance system [1] is an example that utilizes this method. This experimental system is a partial realization of CyberRail, which is an information technology based infrastructure of system concept we are proposing. The system informs appropriate information to passengers via e-mails over cellular phones according to their



travel schedules and current locations. Their locations between starting stations and destination ones are estimated based on their travel schedules and train diagrams. Many users highly rated the service provided by the system, but some people pointed out that guidance at complicated stations is needed.

2.3 Radio Frequency Identification (RFID) tags

There are a lot of varieties of RFID tags, which are different in ranges of communications with receivers and information rates transmitted.

Passive type RFID tags that do not contain batteries can communicate in a range from about a few centimeters through a few ten centimeters. So if a tag is detected by a receiver, it is certain that the tag passes the position where the receiver exists. On the other hand, actions to make the tag close to the receiver are required. Guidance system for handicapped people [2] is an example to utilize passive type RFID tags. Passive type RFID tags are embedded under a Braille block. In a guidance system for visually impaired people, for example, a white cane receives location data from an RFID tag by using an antenna installed in its top, and transmits them to a portable terminal. The portable terminal provides appropriate information to the user.

As active type RFID tags that contain batteries can communicate in a range of a few ten meters, they are effective in cases where the precision of location is required to be about a few meters without any action of users. So this technology was used by an experimental information provision system at shopping mall.

3 Information provision system for passengers

If locations of passengers can be detected at such precision as “they are around a gate” or “they are in a car No.1,” a lot of satisfiable functions of information provision systems will be provided to passengers. As active type RFID tags are considered to be suitable for this precision, we tentatively produced an information provision system by improving the above-mentioned experimental information delivering system for shopping streets.

3.1 Overview of information provision system

The information provision system consists of active type RFID tags, receivers, collection units, location servers, and information delivering servers.

3.1.1 Active type RFID tags

Figure 1(a) shows an active type RFID tag which a passenger receiving guidance information holds. A transmission frequency is 314.950M Hz and a communications distance is about 20 m. The tag transmits its identification information of 16 bits length every 1 through 7 s. Its size is about 4 cm x 3 cm.



3.1.2 Receiver

Figure 1(b) shows a receiver. A receiver receives identification information which active type RFID tags transmit. A receiver has two antennas, one is used to receive identification information, and another is to transmit the receiving identification information of active type RFID tags together with its own identification information to a collection unit via wireless networks dynamically configured among receivers.



Figure 1: (a) Active type RFID tag and (b) Receiver.

3.1.3 Collection unit

A collection unit gathers identification information of both active type RFID tags and receivers transmitted from all receivers, and transmits them to a location server via a serial port. An appearance of a collection unit is the same as that of a receiver except that it has only one antenna.

3.1.4 Location server

A location server identifies locations of active type RFID tags by calculating mean values of coordinates of receivers which receive the identification information of the tags, and notices the locations information to an information delivering server. Figure 2 illustrates an example of a measurement when a person with an active type RFID tag, whose transmission period was 3 s, walked straight at a speed of 1 m/s between the two parallel lines, on which receivers, whose sensitivity was set at about 10 m, were located every 10 m on two lines. Black circles in Fig. 2 are calculated locations by active type RFID tags, and white circles connected with black circles by segments are corresponding real locations.

3.1.5 Information delivering server

An information delivering server traces locations of passengers and sends appropriate e-mails according to their profiles to cellular phones of the passengers at appropriate time.



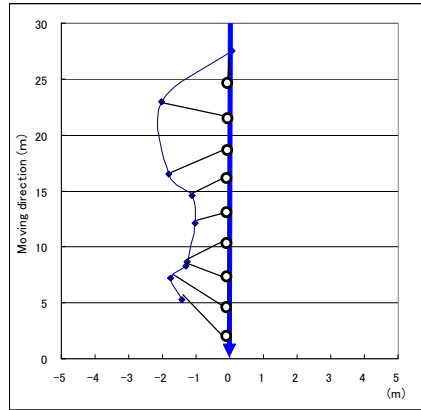


Figure 2: Example of precision of locations by active type RFID tag.

3.2 Items improved from the system for shopping street

The information provision system for passengers and the information delivering system for shopping streets are the same in the sense that both systems deliver information according to current locations and profiles of users. The information provision system for passengers has to take into accounts the following matters or requirements.

- (1) Users of the system move not only in geographically continuous area but also across heterogeneous environments such as from a station to a train. Moreover, trains themselves which are environments for the user move.
- (2) As it is not always possible to communicate between trains and on-ground systems continuously, locations of on-board users are not always grasped.
- (3) It is necessary to take into consideration absolute locations of users as well as relative locations between users and trains. It is also necessary to judge situations such as “in a station” and “on-board” and histories of changes of the situations of users.
- (4) Ticketing information is required as a part of users’ profile information.

To cope with the above matters or requirements, we have improved location servers and information delivering servers as follows.

3.2.1 Location server

It is necessary to trace locations of on-board passengers in order to detect that they get on trains or to notify them of their mistakes if they are wrong vehicles. It is also necessary to define on-board coordinate system apart from coordinates system of stations, and it is desirable to define a coordinate system of each station separately. So we have developed on-board location servers and stations’ location servers, which independently calculate locations of active type RFID tags. Stations’ location servers have the same functions as location servers of the



experimental information delivering system for shopping streets. On-board location servers store calculated locations information of on-board active type RFID tags if communications links between trains and wayside base stations are disconnected, and transmit the stored information to an information delivering server in the lump when the communications links are established. As a result we can continuously gather locations information of passengers and manage histories of their movement including on-board, though some time delays occur in grasping the location information of on-board passengers.

Receivers are also installed appropriately both on-board and in stations. Active type RFID tags are also installed on-board to make stations' location server detect their locations.

3.2.2 Information delivering server

We have provided two basic guidance functions, namely, "guidance on a platform before getting on a train" and "guidance on a platform after getting off a train." An information delivering server has to trace status of passengers in order to deliver necessary guidance information. We have defined three states in this experimental system, namely, "Initial (Outside of railway premise)", "In a station", and "On-board." Figure 3 is a state transition diagram. An information delivering server determines timings of information delivery according to both state transitions of passengers and locations of trains.

We also provided a management function to input and update information about trains such as identification information of active type RFID tags installed on the trains and types of the trains, and profile of passengers such as identification information of active type RFID tags, tickets they hold and their e-mail addresses.

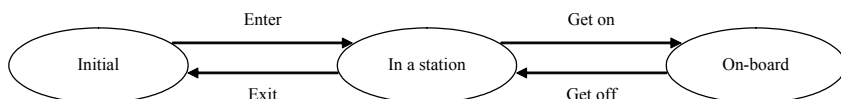


Figure 3: State transition diagram.

3.3 Experiments in compound of RTRI

We carried out experiments to verify functions of our experimental system with a train and mock stations.

3.3.1 Configuration of experiments

Figure 4 shows configurations of experiments. There are two mock stations, namely Station A and Station B, and Local Area Networks (LAN) of the stations are connected by Wireless LAN (WLAN). When a train reaches one of the stations, a communications link between the train and an access point of WLAN of the station is established. There are location servers and receivers at the stations and in the train. An information delivering server is connected to the LAN of Station A.



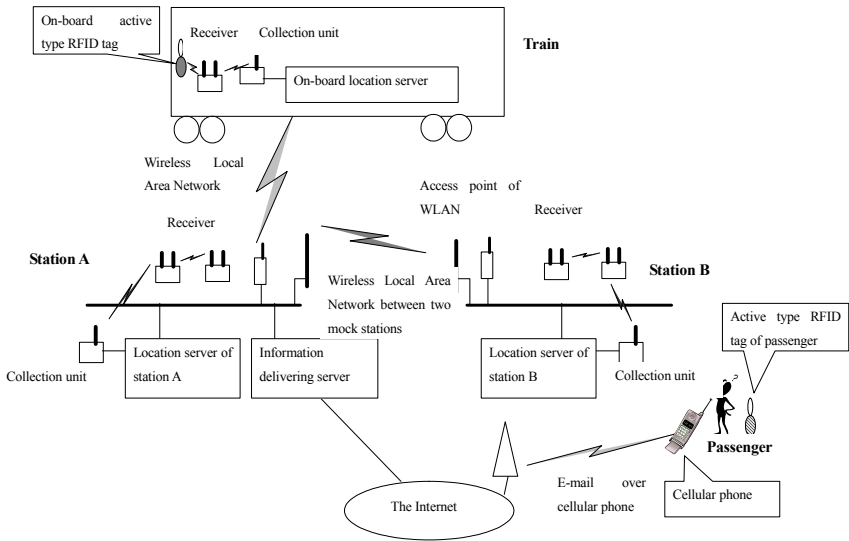


Figure 4: Configurations of experiments.

3.3.2 Scenarios of experiments

We verified operations of the system when users with active type RFID tags and cellular phones moved in the following scenarios.

- (1) A passenger gets on the train at Station A and gets off at Station B. There are two sub-cases, that is, the passenger comes to Station A faster than the train and vice versa. We verified that “guidance on a platform before getting on a train” and “guidance on a platform after getting off a train” were performed at Station A and Station B, respectively.
- (2) A passenger gets on the train at Station A and the train moves to Station B. The passenger gets off the train and gets on the train again at Station B, and the train comes back to Station A, and finally the passenger gets off the train. This simulates the movement of the passenger from Station A to another station via Station B. We verified that “guidance on a platform before getting on a train” was performed when the passenger got on the train at Station A and “guidance on a platform after getting off a train” was performed when the passenger got off the train at Station A, as well as no guidance was performed at Station B.
- (3) There are two passengers P and Q at Station A. P has a ticket for the train X and Q has a ticket for the train Y. We verified that “guidance on a platform before getting on a train” was performed to P but not to Q when train X arrived at the station.

There was almost no delay between the time when the information delivering server transmitted e-mail and the time when the cellular phone received the e-mail. There is possibility of delayed or even undelivered emails, however, as the system cannot control transmission qualities for the Internet and cellular networks.



3.4 Extensions of the experimental system

We extended functions of the experimental system to verify that it could comply with more complicated situations.

3.4.1 Supposed situation

We supposed the situations in which a train was composed of different types of cars. Figure 5 shows an example of such situations. For passengers who get on a train which splits at an intermediate station, guidance information shall be provided according to their destination stations and current locations. We assume that a passenger P travels from station A to Station D and a passenger Q from station A to station C.

- (1) At Station A: P shall be guided “Please go to the platform and stay near Car No.1,” but Q may be guided “Please go to the Platform.”
- (2) On-board: P shall be guided “Please move to Car No.1” if he/she is in Car No.2.
- (3) At Station B: Q shall be guided “You must get off at the next station.”

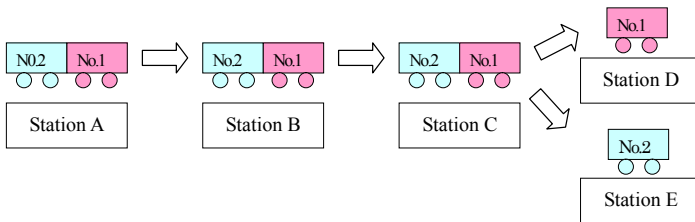


Figure 5: Split train.

3.4.2 Modifications of information delivering server

To cope with the above situations, we modified programs of information delivering server.

- (1) A state “In a station” was divided to two states, “At a gate” and “On a platform.”
- (2) Trains information was extended to include their compositions.

We verified that this system could flexibly cope with sub-divisions of states and additions of contents of guidance.

4 Abnormal situations detection and notification system for railway staff

When alarms are transmitted to the railway staff nearest to the location where any accident happens, required precision of locations of the staffs is about a few meters, and active type RFID tags are suitable. On the other hand, staffs move around not only in station buildings but also outdoors, where receivers of the tags are difficult to install. So it is necessary to use other methods to detect locations. As GPS is such a popular method, we are considering combinations of active type RFID tags and GPS.



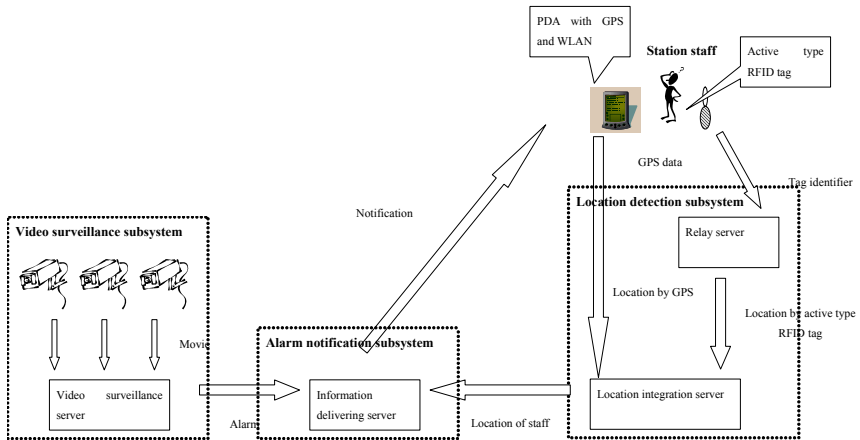


Figure 6: Outline of emergency detection and notification system.

Figure 6 shows an outline of an abnormal situations detection and notification system. This system consists of three subsystems, that is, a video surveillance subsystem, location detection subsystem, and alarm notification subsystem. A station staff holds an active type RFID tag and a Personal Digital Assistance (PDA) with GPS receiver and WLAN interface.

4.1 Video surveillance subsystem

This subsystem consists of video cameras and a video surveillance server. The server detects abnormal situations based on movies transmitted from the video cameras processed by image processing technology, and transmits an alarm which includes a type of the event and occurrence time of the event to an alarm notification subsystem. Examples of the types of events are people gathering, suspicious loitering, left luggage, people fighting, and people fall.

4.2 Location detection subsystem

This subsystem consists of a relay server and a location integration server. The relay server, which is installed at each floor of a station, gathers identification information of active type RFID tags and transmits them to a location gathering server. A PDA periodically calculates its location by its GPS receiver and transmits the result to the location integration server via WLAN. The location integration server finally calculates locations of the PDA from location information by an active type RFID tag and GPS, and transmits the result to an alarm notification subsystem.

4.3 Alarm notification subsystem

This subsystem consists of an alarm notification server. When an alarm notification server receives an alarm from the video surveillance subsystem, it



determines station staffs to whom notifications should be sent considering their current locations, and transmits them via WLAN.

We are now developing the prototype system and will carry out some experiments in our Research Institute. (This work is financially supported in part by the Japanese Ministry of Land, Infrastructure and Transport.)

5 Conclusion

Locations information of persons are essential for ubiquitous services in railways. In this paper, we have proposed two systems, one is to provide information to passengers and another is to detect and alarm abnormal events, both of which utilize locations information detected by active type RFID tags. The information provision system was experimentally manufactured and verified to provide appropriate guidance to passengers.

Next, we will manufacture experimentally the abnormal events detection and notification system and evaluate the effectiveness of the system.

References

- [1] Ogino, T., Tsuchiya, R. : Realization of CyberRail Partial Functions in Actual Environment, Quarterly Report of RTRI, pp.18-22, Vol.46, No.1, 2005
- [2] Matsubara, H., Fukazawa, N., Myojo, S., Tsuchiya, R. : Guidance System for Handicapped People, Quarterly Report of RTRI, pp.13-17, Vol.46, No.1, 2005



A survey on SNCF decision support system tools to supervise and to pilot train traffic in operation

D. Gauyacq

*SNCF Direction de la Recherche et de la Technologie,
Unité de Recherche "Génie Décisionnel Appliqué", Paris, France*

Abstract

In recent years, the SNCF Research Department has developed five different tools to monitor production operations. This supervision has the potential to affect the whole production process. This article will present the role played by the supervisory function in the production process and explain how this function contributes towards making the production process more efficient. In the second part, we shall be describing the basic principles of the tools used for supervisory purposes, both as regards the information systems that support them, the alarm, browse and display functions used for the production data, and the functions supporting the decision-making process. In the rest of the text, we shall be illustrating these principles by providing a brief description of two of the specific tools in use: GEDEON, for freight operations and EXCALIBUR, for overall management of trains in circulation. In conclusion, we shall attempt to show the work necessary to make these tools more efficient, whilst recognising that they can only be properly effective if associated with new forms of production organisation better adapted to the target purpose.

Keywords: supervision, operational management of traffic, man and machine interfaces, optimisation, quality of service, production system.

1 Introduction

Compared to most of the other existing transport modes, the railway can boast a large number of intrinsic advantages in terms of sustainable development. This is particularly true in the case of long-distance bulk traffic, an area where the energy required per unit carried is at its minimum. Yet, in a competitive



environment with a market economy, this factor alone is not enough to make rail the obvious choice. In Europe, to enhance the future prospects of the rail mode, freight in particular, it is first and foremost necessary to boost production system efficiency. For this, substantial and quantifiable progress needs to be made to drive production costs radically down and vastly improve customer service standards. It is reasonable to conjecture that the action to be taken on these two fronts will be more complementary than contradictory, as we shall attempt to demonstrate in section 2.2.

This article will have the following structures: in the first part, we shall run briefly through the role played by the supervisory function in the production process. We shall be explaining how this function contributes towards making the production process more efficient. In the second part, we shall be describing the basic principles of the tools used for supervisory purposes, both as regards the information systems that support them, the alarm, browse and display functions used for the production data, and the functions supporting the decision-making process. In the rest of the text, we shall be illustrating these principles by providing a brief description of two of the specific tools in use: GEDEON, for freight operations and EXCALIBUR, for overall management of trains in circulation. In conclusion, we shall attempt to show the work necessary to make these tools more efficient, whilst recognising that they can only be properly effective if associated with new forms of production organisation better adapted to the target purpose.

2 Supervision of production: a key factor in the industrialisation process

2.1 Role assigned to supervision

The functional objectives associated with rail system production control are relatively easy to list, since the process consists of: (i) monitoring the execution of a production plan; (ii) checking that it is properly performed; (iii) reporting any disparities noted to an operator. Then he must, where relevant, anticipate other knock-on effects, in order to contain the consequences on the value of the service produced, in both quality and quantity terms. This process is in no way automatic, for the operator is positioned at the critical junction between observation of a problem and remedial action. (Actually, he acts in accordance with the principle of limited rationality as defined by H. Simon [5].) In the action he takes, the supervisor will use his professional experience, his knowledge of how the production process works and the margins for manoeuvre offered him by the system. This makes him a crucial player in the production process. (This need for a relatively impartial system of monitoring is virtually synonymous with the process of industrialisation. A good illustration of this is the “*silent controller*” introduced by Robert Owen [4] in the early 19th century at his *New Larnak* factory. That innovation was theorised by Jeremy Bentham [1]. It is this same idea that was developed by Japanese engineer Taiichi Ohno (Toyota



adviser) for whom “*visual management*” was vital [2].) It also means that he, too, needs to be supervised as part of the production process.

2.2 Contribution of supervision to improving railway operating efficiency

The rail network is a system based on economies of scale. (But is also a system of “*economy of system*” where productivity can be considerably enhanced without major investment but by improving the operational control/command system, primarily in order to co-ordinate the interaction between sub-systems and production players [3].) To drive down costs, it is therefore first necessary to make maximum use of railway infrastructure to reduce the unit operating cost, which in one way or another is always ultimately charged to production. Next, it is necessary to enhance the productivity of the main components of the production process, in other words the locomotives, wagons or coaches and workforce (including drivers). In fact, to exert proper control over production operations and thereby guarantee efficiency and quality, it is necessary to produce a formal definition of these operations. This definition, and above all its application, should not be allowed to have a diminishing effect or make it impossible for workers to take the initiatives necessary to meet the fundamental quality goals. It goes without saying that these different methods for enhancing system efficiency are mutually interdependent and part of an interrelated virtuous circle. Progress made in terms of lower costs and /or higher quality is bound to boost demand and thereby drive down fixed unit costs. Properly controlled and optimised technical production operations strictly applied should reduce the number of production incidents. In return this will make it possible to make greater use of production resources by lessening the unproductive margins set aside to allow for unexpected events, themselves often the direct result of improvised adaptation.

3 Basic principle behind the supervisory tools developed at SNCF

In recent years, the SNCF Research Department has developed five different tools to monitor production operations. This supervision has the potential to affect the whole of the production process: from the formation of trains in stations (AGORA), operation of TGV high-speed trains (COLT), suburban trains (ECLER), freight trains (GEDEON) but also to management at national level by the infrastructure manager of all these different movements (EXCALIBUR). The principles behind these tools are virtually identical in each case. Figure 1, shows in diagrammatic form the functional architecture we shall be describing in more precisely.

The main function of these tools is to monitor the actual implementation of the production plans and alert the operators as soon as any notable slippages appear.



Each is based on four categories of information:

- 1) The production plan to be implemented, for example, the timetable diagram showing the trains to be worked, the routes they are to follow and the time at which they are expected to pass certain specific points;
- 2) The resources allocated to achieve this production plan: locomotives and driver rosters and, in addition, details of stopping (at platforms) and side-tracking;
- 3) Monitoring the production process via data showing, for example, the progress of trains over the network. This information comes mainly from fixed sensors installed in the track or from systems on board the vehicles themselves which enable their position to be established by GPS and transmitted by GSM;
- 4) The rules whereby it is possible to detect any slippage in the timing of the different operations in the production process. This can be seen or deduced, for example, when information expected fails to arrive.

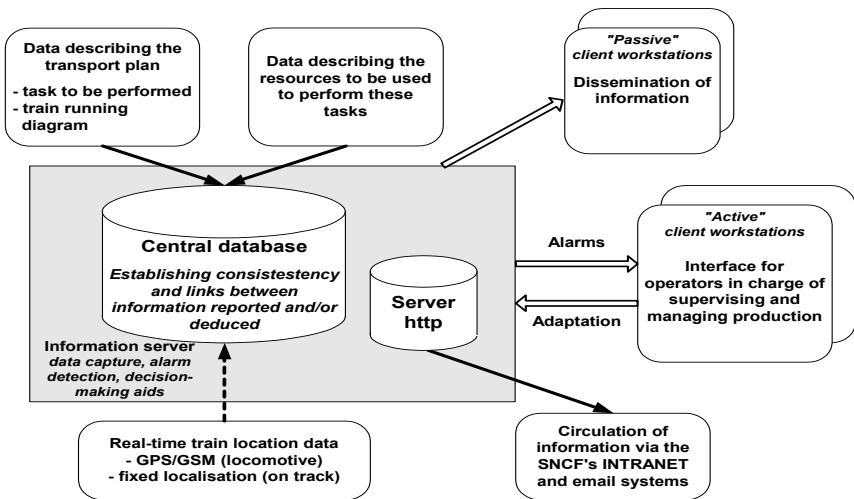


Figure 1.

All these tools are fed by information systems that existed before to their creation. In particular, we exploit the systems used to produce the different production plans for the specific types of traffic (passengers, freight, etc.) and/or specific resources (locomotives, driver, etc.). All information is periodically collected, generally on a daily basis, and input into a central database specific to each tool. Technically, all these tools are structured on a client/server basis so that data can be managed centrally and connected up to a large number of client (operator) workstations spread over a wide area, depending on the production requirements and organisation adopted. For a given application, this means distributing different services and information to the workstations connected to the network in relation to the different roles and functions of the operators working on them.



To help operator in his task, the tool reports those production slippages that have reached a given threshold α_i in relation to category i of the operation monitored, for example, when a train is still not reported as having passed a particular point $\alpha_{\text{train.delay}}$ minutes after the expected time. This threshold value is a parameter that the operator can adapt to circumstances. To make it easier for the operator to understand the problems, ergonomic interfaces display the potential consequences of the initial problem. For example, if a train is running late, this could mean that a driver travelling to work on this train would arrive late to take up his duty on train T_x . The operator will have to work out a proposal to cater to this eventuality, for example, by switching round two driver rosters or by calling in a reserve driver or by noting the delay and the way it will affect train T_x . In this last case, he should warn the department in charge of operating train T_x to prevent it from being taken by surprise and, if necessary, to enable it to advise customers of the problem. To report problems encountered and decisions made, most of these tools have an *http* server connected to their database. This contains all the information judged useful in HTML format, which can then be polled via the SNCF's *Intranet* network. Access to this information is restricted to authorised personnel only.

The man/machine interfaces rely essentially on:

- ❑ Mapping data, in the form of geographical information systems displaying the physical location the different trains, locomotives, groups of wagons or the location of operating incidents;
- ❑ Graphic diagrams of the “*time/resource*” variety, for example, the rosters of the different locomotives in use. These show the resources used to complete a production task and the next tasks for which these resources are scheduled.
- ❑ Tables listing the alerts given, describing the delays incurred, the resources affected and the possible knock-on effects on other tasks;
- ❑ More detailed descriptions of the operations to be conducted, especially as regards train routes, schedules, driver's relief or locomotive unit transfer points, train composition, etc.
- ❑ Menus and/or tab bars to browse through pre-set screens, each of which may contain specific display and/or data, capture menus.

During the development of these tools, a number of decision-making aid functions were produced that have yet to be fully agreed by users. As a result, these functions are still at the prototype stage and require further enhancement. These support functions relate to three major categories of action which correspond, more or less, to the seriousness of the problem to be addressed and the time it will take for the situation to return more or less to normal:

- ❑ *If/what* tests to establish, more or less accurately, what the ultimate consequences of a particular decision will be on action in progress;
- ❑ Assessment and/or application of predetermined problem-solving scenarios. The operator may perform these actions automatically.
- ❑ Computing functions of the “operational research” type, to adapt the production schedule.



4 GEDEON – FREIGHT production supervision tool

The tools (COLT, ECLER, GEDEON) used to supervise train movements are all based on similar functions. The differences between them lie essentially in the information systems to which they are connected. The most important of these differences relate to the type of traffic under control; other, more minor differences result from the way in which the production service in charge of the traffic is organised. In this paper, we shall simply describe GEDEON, since it is widely used in the different FREIGHT production centres.

4.1 GEDEON functions

The central system supervises all FREIGHT sector trains on a real-time basis. Passenger trains headed by locomotives next scheduled to haul a FREIGHT train are also monitored. The corresponding information is placed at the disposal of the different production supervision centres. Each dedicated centre is considered to be a client of the central system and, as such, has the authority to supervise a series of trains. These latter are selected in line with the FREIGHT operations organisation which, from a supervisory standpoint, may (but does not necessarily always) have three categories of centre dedicated to:

- ❑ a type of traffic, for example, combined transport;
- ❑ a major traffic route, in a bid to maximise flows;
- ❑ a local zone in charge of gathering and delivering wagons.

GEDEON is supplied with the following information:

1. The data corresponding to the production schedule for a given period, for example 24 hours:
 - ❑ The diagram of all FREIGHT train movements, some 2,500 trains each day;
 - ❑ The list of wagon groups to be forwarded in these trains;
2. Data on the resources, locomotives and drivers earmarked for use in this production process:
 - ❑ The references of the locomotives scheduled to haul the trains;
 - ❑ The daily rosters of the drivers;
3. Production monitoring data assimilated to the pattern of train movements over the network:
 - ❑ Information about the time trains are scheduled to pass specific points on the network captured via fixed transponders or input by signalmen;
 - ❑ Automatic locomotive positioning data via GPS/GSM;

When a locomotive or a train is discovered to be running late by a given threshold delay, it is highlighted on the map at its latest known position and flagged up in the alarm tables. If the delay is unlikely to have any effect on the deployment of resources, a yellow symbol is used, otherwise the symbol is red. The resources (locomotive, driver) affected by this delay are also marked with the same colour code plus an indication of the estimated duration of the delay.



4.2 Illustration of man/machine interfaces

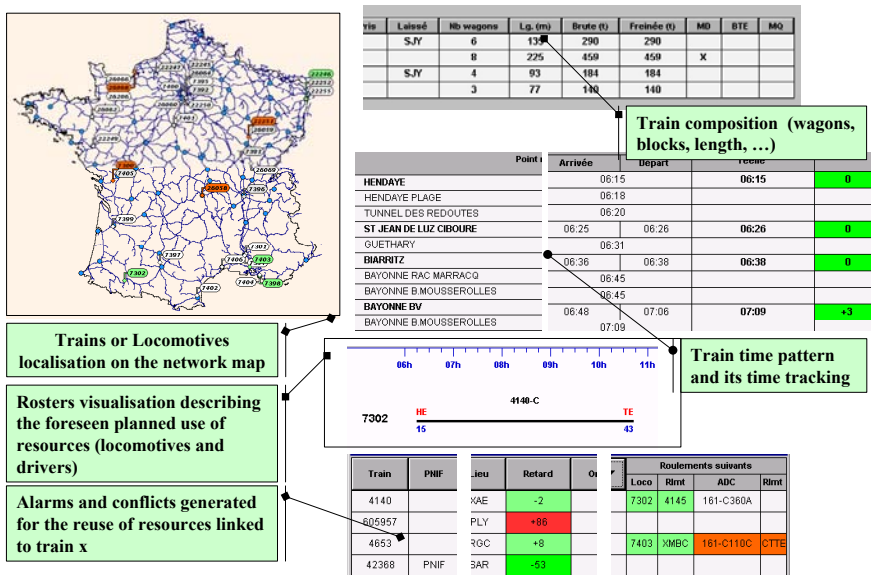


Figure 2: MMI to present information about freight traffic and resources of production.

5 EXCALIBUR – overall train movement supervision tool

EXCALIBUR is a tool that is in use at the SNCF's National Operations Centre (CNO). This centre is responsible for co-ordinating the activities of the 22 regional traffic control centres in charge of managing traffic for the tasks assigned to the infrastructure manager (IM). It is also the premise where the IM and the passenger and freight traffic operators can interact operationally to ensure that the knock-on effects of operating incidents affecting network capacity are managed with maximum efficiency.

5.1 EXCALIBUR functions

The EXCALIBUR system is based on data describing the railway system. This description is geographically referenced, in other words the information may be shown on a skeleton map. The basic information comes from the RESEAU database and is used to attribute different qualities to the railway lines represented. As an example, the following line characteristics may be displayed on request: {type of electrification; zone with/without traffic control; train running supervision system; possible throughput in relation to incidents reported; maximum speed; radio equipment; type of line (high-speed, conventional, narrow gauge); type of block section; operating system; type of traffic for each category (TGV, freight, main-line, Paris suburban)}. The map also shows the national and SNCF administrative boundaries of the territory. It also shows the

network in different degrees of detail depending on the amount of zoom applied. The most critical points in operating terms (tunnels, level crossings) can also be displayed, as can technical documents showing the track diagram of areas with complex layouts.

Any current incidents jeopardising the flow of traffic are marked on the map using symbols to show the degree to which throughput is affected. All the information available to describe the incident (time, place, cause, and consequences) can be accessed via the map references. Information about the development of the incident, action taken, steps to enable operations to continue are described, thus enabling each incident and its consequences to be traced and recorded. This adds up to an invaluable source of information and feedback for calculating the expected duration of future incidents and dynamically adapting these calculations to typical incident management techniques. For example, the knock-on effects of accidents involving people are always serious but the duration of the resulting traffic stoppage can vary widely, from one to several hours. It will be possible to give a more precise idea about the duration of each specific incident from information feedback by the system. This information may indicate whether or not the accident victim is still alive, whether or not the emergency services have reached the scene, whether or not the police or the judiciary officials are involved, etc.

The HOUAT system supplies details of the trains due to operate as part of the day's production schedule. The position data that can be used to follow train movements over the network is transmitted in real time by the BREHAT system. This information is supplemented by data input by EXCALIBUR users, who enter details of each incident, its repercussions and its development. This information is accessible to a whole series of duly authorised persons via Intranet (some 300 regularly connected users, when they need it). When an incident looks like affecting a region not immediately concerned, an email is sent to the traffic controllers in this region advising them to go to the EXCALIBUR site to follow developments.

For the moment EXCALIBUR offers two types of decision-making aid. The first is based on statistics approach and is designed to establish the likely duration of a given type of incident, the standard deviation for the particular incident type. This value is adjusted in relation to information coming in about the actual incident and its progression, action taken and the results observed. Work is at present in hand to introduce this approach on a systematic basis by using historical data on fixed installation failures and by finding the best way of grouping together and segmenting this data to establish the most accurate and reliable forecasts. The second relates to managing re-routing operations when major incidents prevent trains from running on certain sections of line. For the moment (cf. illustration) the support provided is more than information enabling the operator to:

- ❑ identify the trains to be re-routed and those already scheduled to run over alternative routes;
- ❑ identify the cost of using a particular alternative route (extension of journey time, length, etc.);



- ❑ select the right way to manage the incident (cancelling a train, maintaining it on the original route, diverting it, etc).

Ultimately this information will be extended to include action proposals suggested by an optimisation module that will (re)schedule current and future trains by attributing train paths and routes to minimise the consequences of the particular incident.

5.2 Illustration through Man Machine Interfaces (MMI)

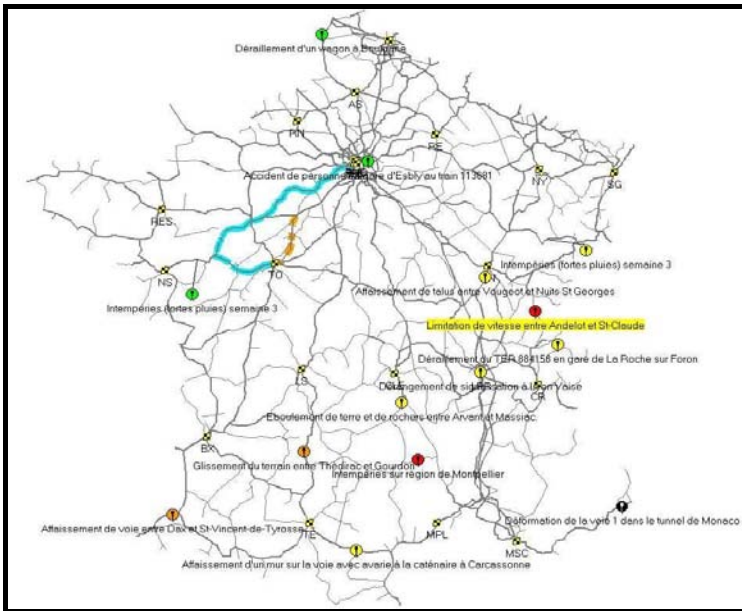


Figure 3: MMI to visualise infrastructure problems (colour circle) and rerouting possibilities (blue line) when a section line is blocked (dash yellow line).

6 Conclusion and prospects

The supervision tools presented above form sound foundations on which new projects in two complementary directions will be developed. The first of these involves improving production standards by tightening controls and attaching importance to performances achieved. Obviously, this can only be done if the production system is first standardised (as in the AGORA project) and by monitoring the action of those involved in the production process. (AGORA serves to supervise (i) the preparation of trains and (ii) the flow of passengers alighting from arriving or boarding departing trains. Its main role is to ensure that the instructions contained in the quality manual are applied in practice. For example, a train (in its correct formation, cleaned, etc.) should be at the platform α , z minutes before its departure time t_0 ; that the driver should report for duty on



the train x minutes before t_0 , and so on.) Where this latter point is concerned, the performance assessments (differences in relation to standards set and production results) allowed by these tools could be carried out for each of the production control centres, with the result that this form of accountability could play a dual role:

- ❑ checking that the various tasks are performed in line with the official procedures, that remedial action (with the help of tools in the case of complex procedures) is taken sufficiently quickly and thoroughly;
- ❑ Appreciation of the work done and the contribution of the different teams and individuals to ensuring that the overall system functions smoothly.

It goes without saying that this type of action should be part of a renewal of organisational and managerial system for it to be fully effective.

The second line of approach is more technical and would consist of introducing more efficient decision-making functionalities:

- ❑ to be able better to anticipate the consequences of operating problems and incidents and provide information facilitating the co-ordination of the production players and, by extension, the customer (by making the first such functions systematic in EXCALIBUR);
- ❑ to adapt the production plans as they are being carried out on a semi-automatic basis when, for whatever reason, it is no longer possible to stick to the original plan.

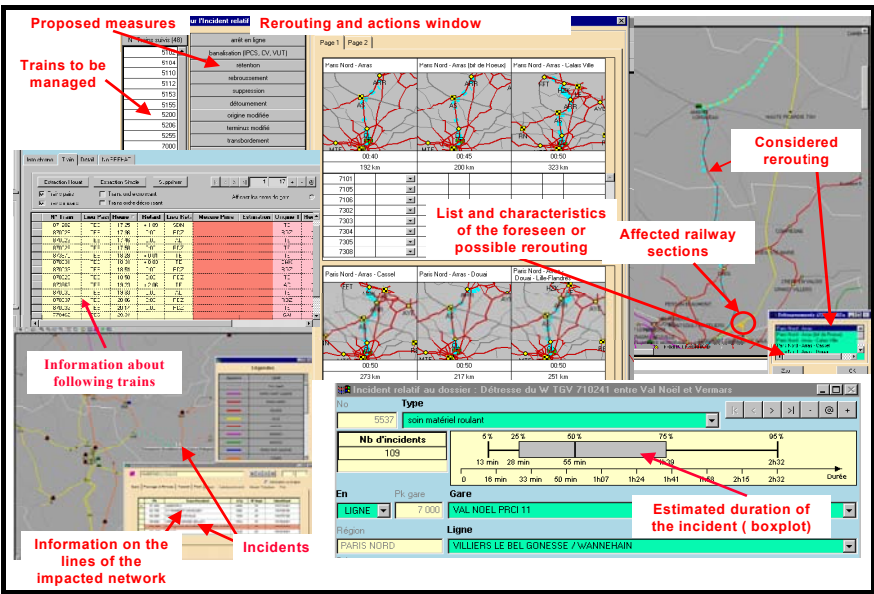


Figure 4: MMI to manage rerouting of trains when a line section is blocked.

In both cases, the extra functions that need to be developed or enhanced can only provide support and assistance in actions that involve changing (or at least stimulating and boosting) the existing production organisation.

In our opinion, the main contribution of these tools and the steps planned to build up and extend their supervisory functions lies in making staff recognise the need for change and motivating them to improve their production standards. The knock-on effect is to enhance productivity through tighter control of operations. Conversely, for these tools to work efficiently they need to be used by staff with the requisite new skills and, above all, the ability to understand the problem-solving logic involved so they can properly interpret and exploit their new resources.

References

- [1] Jeremy Bentham *Optique The Panopticon Writings*, Edition Mille et Une Nuits, No. 398, 2002.
- [2] Benjamin Coriat, “*Penser à l’envers, Travail et organisation dans l’Entreprise japonaise*” (Thinking backwards and Japanese company organisation), Christian Bourgeois Publisher, 1991.
- [3] Bernard Lacroix *Technical Systems approach applied to the Air Traffic Management: the central role of the economic forces*”, Conference Delft, 2001, www.delft2001.tudelft.nl/paper%20files/paper1128.doc
- [4] R. Owen, “*Textes choisis*” (Selected texts), p 79, Editions Sociales, 1963
- [5] H.A. Simon, J.-G. Marches *organisations*” (Organisations), DUNOD, 1999.



This page intentionally left blank

Section 5

Computer techniques

This page intentionally left blank

Distributed constraint satisfaction problems to model railway scheduling problems

P. Tormos², M. Abril¹, M. A. Salido¹, F. Barber¹, L. Ingolotti¹
& A. Lova²

¹*DSIC, Universidad Politécnica de Valencia, Spain*

²*DEIOAC, Universidad Politécnica de Valencia, Spain*

Abstract

Railway Scheduling is considered to be a difficult and time-consuming task. Despite real railway networks being modelled as Constraint Satisfaction Problems (CSPs), they require a huge number of variables and constraints. The general CSP is known to be NP-complete; however, distributed models may reduce the exponential complexity by dividing the problem into a set of subproblems. In this work, we present several proposals to distribute the railway scheduling problem into a set of sub-problems as independently as possible. The first technique carries out a partition over the constraint network, meanwhile the second distributes the problem by trains and the third technique divides the problem by means of contiguous stations.

1 Introduction

The literature of the 1960s, 1970s, and 1980s relating to rail optimization was relatively limited. Compared to the airline and bus industries, optimization was generally overlooked in favor of simulation or heuristic-based methods. However, Cordeau et al. [1] point out greater competition, privatization, deregulation, and increasing computer speed as reasons for the more prevalent use of optimization techniques in the railway industry. Our review of the methods and models that have been published indicates that the majority of authors use models that are based on the Periodic Event Scheduling Problem (PESP) introduced by Serafini and Ukovich [2]. The PESP considers the problem of scheduling as a set of periodically recurring events under periodic time-window constraints. The model generates disjunctive constraints that may cause the exponential growth of the computational



complexity of the problem depending on its size. Schrijver and Steenbeek [3] have developed CADANS, a constraint programming- based algorithm to find a feasible timetable for a set of PESP constraints. The scenario considered by this tool is different from the scenario that we used; therefore, the results are not easily comparable. Nachtigall and Voget [4] also use PESP constraints to model the cyclic behavior of timetables and to consider the minimization of passenger waiting times as the objective function. Their solving procedure starts with a solution that is obtained in a way similar to the one that timetable designers in railway companies use. This initial timetable is then improved by using a genetic algorithm. In our problem, the waiting time for connections is not taken into account because we only consider the timetabling optimization for a single railway line.

The train scheduling problem can also be modeled as a special case of the job-shop scheduling problem (Silva de Oliveira [5], Walker and Ryan [6]), where train trips are considered as jobs that are scheduled on tracks that are regarded as resources. The majority of these works consider the scheduling of new trains on an empty network. However, railway companies usually also require the optimization of new trains on a line where many trains are already in circulation (that is, trains that have a fixed timetable). With this main objective, Lova et al. [7] propose a scheduling method based on reference stations where the priority of trains, in the case of conflict, changes from one iteration to another during the solving process.

Our goal is to model the railway scheduling problem as a Constraint Satisfaction Problems (CSPs) and solve it using constraint programming techniques. However, due to the huge number of variables and constraints that this problem generates, a distributed model is developed to distribute the resultant CSP into a semi-independent subproblems such as the solution can be found efficiently.

In parallel computing, many researchers are working on graph partitioning [8], [9]. The main objective of these techniques is to divide the graph into a set of regions such that each region has roughly the same number of nodes and the sum of all edges connecting different regions is minimized. Thus, we can use ideas about graph partitioning, when dealing with railway scheduling problem, to distribute the problem into a set of sub-problems.

In this work, we propose several ways to distribute the railway scheduling problem. It is partitioned into a set of subproblems by means of graph partitioning, by means of types of trains and by means of contiguous constraints.

2 Problem modelling: a constraint satisfaction problem

A CSP consists of a set of variables $X = \{x_1, x_2, \dots, x_n\}$, each variable $x_i \in X$ has a set D_i of possible values (its domain), and a finite collection of constraints $C = \{c_1, c_2, \dots, c_p\}$ restricting the values that the variables can simultaneously take. A solution to a CSP is an assignment of values to all the variables so that all constraints are satisfied; a problem is *satisfiable* or *consistent* when it has a solution at least. A partition of a set C is a set of disjoint subsets of C whose union is C . The subsets are called the blocks of the partition.



A running map contains information regarding railway topology (stations, tracks, distances between stations, traffic control features, etc.) and the schedules of the trains that use this topology (arrival and departure times of trains at each station, frequency, stops, crossings, etc.).

A distributed CSP (DCSP) is a CSP in which the variables and constraints are distributed among automated agents [10]. Each agent has some variables and attempts to determine their values. However, there are interagent constraints and the value assignment must satisfy these interagent constraints. In our model, there are k agents $1, 2, \dots, k$. Each agent knows a set of constraints and the domains of variables involved in these constraints.

2.1 Constraints in the railway scheduling problem

There are three groups of scheduling rules in our railway scheduling problem: traffic rules, user requirements rules and topological rules. A valid running map must satisfy the above rules. These scheduling rules can be modelled using the following constraints, where variable $TA_{i,k}$ represents that train i arrives at station k and the variable $TD_{i,k}$ means that train i departs from station k :

1. Traffic rules guarantee crossing and overtaking operations. The main constraints to take into account are:
 - *Crossing constraint*: Any two trains going in opposite directions must not simultaneously use the same one-way track.
 - *Overtaking constraint*: Two trains going at different speeds in the same direction can only overtake each other at stations.

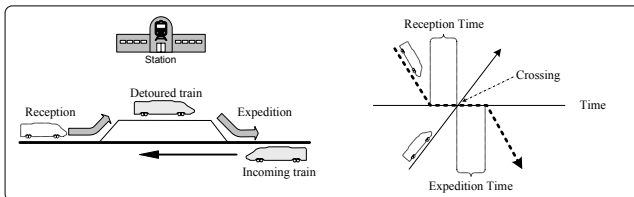


Figure 1: Constraints related to crossing and overtaking in stations.

- *Expedition (Reception) time constraint*. There exists a given time to put (detour) a train back on (from) the main track so that crossing or overtaking can be performed.
2. User Requirements: The main constraints due to user requirements are:
 - *Type and Number of trains* going in each direction to be scheduled.
 - *Path of trains*: Locations used and *Stop time* for commercial purposes in each direction.
 - *Scheduling frequency*. Train departure must satisfy frequency requirements in both directions.



- *Departure interval* for the departure of the first trains going in both the up and down directions.
3. Railway infrastructure topology and type of trains to be scheduled give rise to other constraints to be taken into account. Some of them are:
- *Number of tracks in stations* (to perform technical and/or commercial operations) and the number of tracks between two locations (one-way or two-way).
 - *Time constraints*, between each two contiguous stations,

The complete set of constraints, including an objective function, transform the CSP into a constraint satisfaction and optimization problem (CSOP), where the main objective function is to minimize the journey time of all trains. Variables are frequencies and arrival and departure times of trains at stations. The complete CSOP is presented in Figure 2. Let us suppose a railway network with r stations, n trains running in the down direction, and m trains running in the up direction. We assume that two connected stations have only one line connecting them. $Time_{i,k-(k+1)}$ is the journey time of train i to travel from station k to $k+1$; $TS_{i,k}$ and $CS_{i,k}$ represent the technical and commercial stop times of train i in station k , respectively; and ET_i and RT_i are the expedition and reception time of train i , respectively.

3 Partition proposals

3.1 Partition proposal 1: graph partitioning

The first way to distribute the problem is carried out by means of a graph partitioning software called METIS [11], for the purpose of this distribution, the model constraints are converted into binary constraints. METIS provides two programs *pmetis* and *kmetis* for partitioning an unstructured graph into k equal size parts. However, this software does not take into account additional information about the railway infrastructure or the type of trains to guide the partition, so the generated clusters may not be the most appropriate and the results are not appropriate.

3.2 Partition proposal 2: by trains

The second model is based on distributing the original railway problem by means of train type. Each agent is committed to assign values to variables regarding a train or trains to minimize the journey travel. Depending on the selected number of partitions, each agent will manage one or more trains. Figure 3a shows a running map with 20 partition, each agent manages one train. This partition model has two important advantages: Firstly, this model allow us to improve privacy.

In this way, the partition model gives us the possibility of partition the problem such as each agent is committed to a operator.

This model allow us to manage efficiently priorities between different types of trains (regional trains, high speed trains, freight trains).



Integer Variables $\forall i = 1..n \forall k = 1..r$
$TA_{i,k}$ 0..86400;
$TD_{i,k}$ 0..86400;
Objective Function
(1) $\text{Min } \sum_{i=1}^{i=m} (TA_{i,r} - TD_{i,1}) + \sum_{j=1}^{j=m} (TA_{j,1} - TD_{j,r});$
Subject to
Frequency Constraints $\forall i = 1..n \forall k = 1..r$
(2) $TD_{i+1,k} - TD_{i,k} = \text{Frequency};$
Time Constraints $\forall i = 1..n \forall k = 1..r$
(3.1) $TA_{i,k+1} - TD_{i,k} = \text{Time}_{i,k-(k+1)};$
(3.2) $TA_{i,k} - TD_{i,k+1} = \text{Time}_{i,k-(k+1)};$
Stations Time Constraints $\forall i = 1..n \forall k = 1..r$
(4) $TD_{i,k} - TA_{i,k} - TS_{i,k} = CS_{i,k};$
Constraints to limit journey time $\forall i = 1..n \forall j = 1..m$
(5.1) $TA_{i,r} - TD_{i,1} \leq (1 + \frac{\delta}{100}) * \text{Time}_{i,1-r};$
(5.2) $TA_{j,1} - TD_{j,r} \leq (1 + \frac{\delta}{100}) * \text{Time}_{j,r-1};$
Crossing Constraints $\forall i = 1..n \forall j = 1..m \forall k = 1..r$
(6) $TA_{i,k+1} < TD_{j,k+1} \vee TA_{j,k} < TD_{i,k};$
Overtaking Constraints $\forall i = 1..n \forall j = 1..m \forall k = 1..r$
(7) $TD_{i,k} < TD_{j,k} \rightarrow TA_{i,k+1} < TA_{j,k+1};$
Expedition time Constraints $\forall i = 1..n \forall j = 1..m \forall k = 1..r$
(8) $TA_{i,k} + ET_j \leq TD_{i,k} \vee TA_{j,k} + ET_i \leq TD_{j,k};$
Reception time Constraints $\forall i = 1..n \forall j = 1..m \forall k = 1..r$
(9) $TA_{i,k} + RT_i \leq TA_{j,k} \vee TA_{j,k} + RT_j \leq TA_{i,k};$

Figure 2: Formal model of the railway scheduling problem.

3.3 Partition proposal 3: by stations

The third model is based on distributing the original railway problem by means of contiguous stations. Therefore, a logical partition of the railway network can be carried out by means on regions (contiguous stations). To carry out this type of partition, it is important to analyze the railway infrastructure and detect restricted regions (bottlenecks). To balance the problem, each agent is committed to a different number of stations.

Figure 3b (up) shows a running map to be scheduled. The set of stations will be partitioned in block of contiguous stations and a set of agents will coordinate to achieve a global solution (Figure 3b (down)). Thus, we can obtain important results such as railway capacity, consistent timetable, etc.

4 Evaluation

In this section, we carry out an evaluation between our distributed models and a centralized model. This empirical evaluation was carried out over a real railway



infrastructure that joins two important Spanish cities (La Coruña and Vigo). The journey between these two cities is currently divided by 40 stations. In our empirical evaluation, each set of random instances was defined by the 3-tuple $\langle n, s, f \rangle$, where n was the number of periodic trains in each direction, s the number of stations and f the frequency. The problems were randomly generated by modifying these parameters.

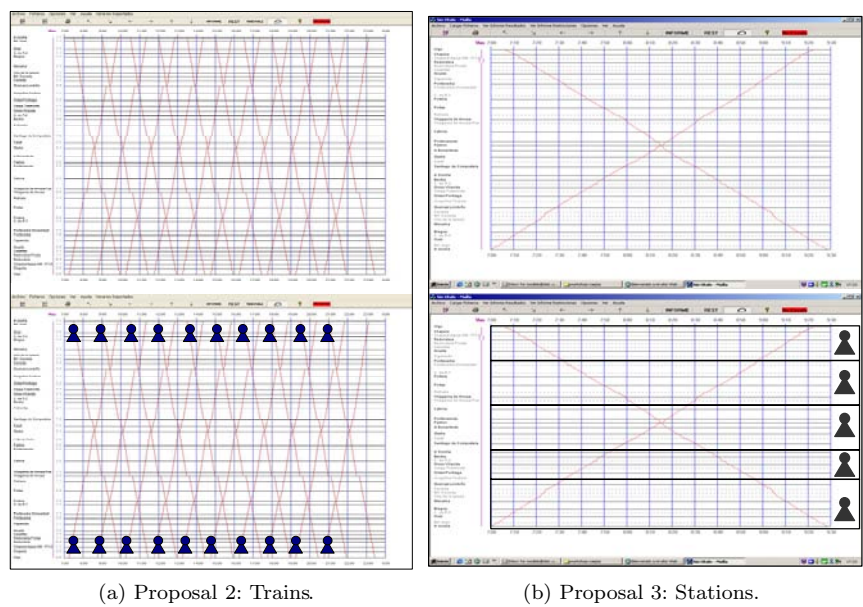


Figure 3: Distributed railway scheduling problem.

Figure 4a shows the running time when the number of trains increase, meanwhile Figure 4b shows the running time when the number of stations increase. In both Figures, the partition model selected was partition proposal 2, where the number of partition/agents was equal to the number of trains. In both figures, we can observe that the running time increased when the number of trains increased (Figure 4a) and when the number of stations increased (Figure 4b). However, in both cases, the distributed model maintained better behavior than the centralized model.

General graph partitioning applications work well in general graphs. However, in the railway scheduling problem, we did not obtain good results using these softwares. We evaluate the partition proposal 1 by using METIS in several instances of the problem. However, the obtained results were even worse in the distributed model than in the centralized model. We studied the partitions generated by METIS and we observed that the journey of a train is partitioned in several clusters, and each cluster was composed by tracks of trains in opposite

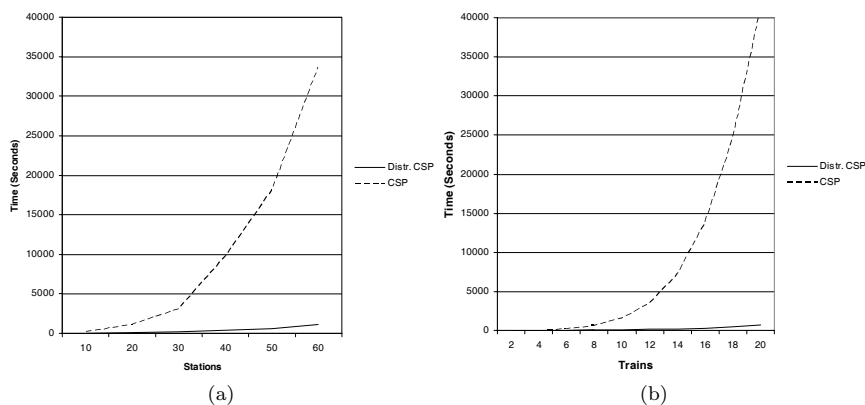


Figure 4: Running Time when the number of trains (a)/stations (b) increased.

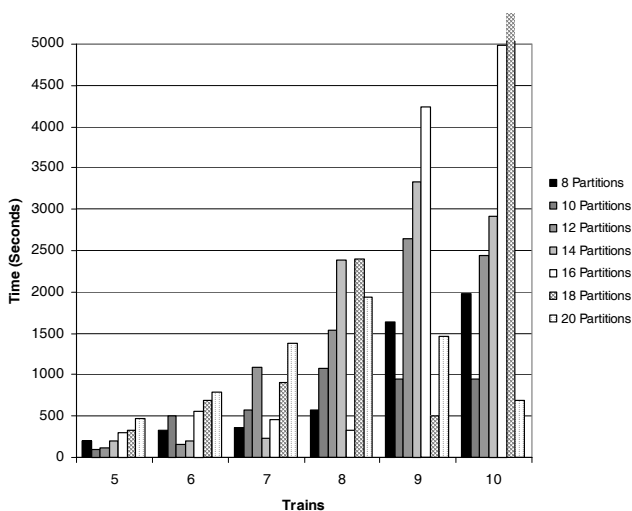


Figure 5: Running Time with different partitions.

directions. This cluster is easy to solve but very difficult to propagate to other agents. Furthermore, the following partition proposals make the contrary, that is, they never join tracks of trains in opposite directions.

So, we can conclude that the problem is very dependent of the partition that we carry out, and a general partition based on low connectivity is not always the best solution.

The partition proposal 2 was the best of the partition proposals, where we can schedule many trains in large railway infrastructure. However, how many partitions must divide the railway problem? If we select a large number of partitions, each subproblem is very easy, but the efficiency decreased due to communication



messages. If we select a low number of partitions, each subproblem may be also difficult to solve. So, an appropriate number of partitions must be studied to solve the problem efficiently.

Figure 5 shows the running time with different partitions in problems where we fix the frequency (120 minutes), the number of stations (20) and the number of trains in each direction was increased from 5 to 10. Each instance was solved by the distributed model with different number of partitions (8, 10, 12, 14, 16, 18, 20 partitions). We can observe the direct relation between the number of trains and the number of partitions. Thus, a agent was committed to schedule a train. If a agent was committed to schedule several trains, the efficiency decreased. Similar results happened when a train was scheduled by several agents.

5 Conclusion and future work

In this paper, we present a distributed model for solving the railway scheduling problem, in which several proposals are developed to distribute the railway scheduling into a set of sub-problems as independent as possible. The evaluation section shows the railway scheduling problem can be solved more efficiently in a distributed way.

Acknowledgements

This work has been partially supported by the research projects TIN2004-06354-C02- 01 (Min. de Educación y Ciencia, Spain-FEDER), FOM-70022/T05 (Min. de Fomento, Spain) and FP6-021235-2, IST-STREP (UE). We appreciate the assistance of Jose Estrada (Spanish Administration of Railway Infrastructure - ADIF).

References

- [1] Cordeau, J., Toth, P. & Vigo, D., A survey of optimization models for train routing and scheduling. *Transportation Science*, **32**, pp. 380–446, 1998.
- [2] Serafini, P. & Ukovich, W., A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics*, pp. 550–581, 1989.
- [3] Schrijver, A. & Steenbeek, A., Timetable construction for railned. *Technical Report, CWI, Amsterdam, The Netherlands*, 1994.
- [4] Nachtigall, K. & Voget, S., Minimizing waiting times in integrated fixed interval timetables by upgrading railway tracks. *European Journal of Operational Research*, **103**, pp. 610–627, 1997.



- [5] Silva de Oliveira, E., Solving single-track railway scheduling problem using constraint programming. *Phd Thesis Univ of Leeds, School of Computing*, 2001.
- [6] Walker, S.J., C. & Ryan, D., Simultaneous disruption recovery of a train timetable and crew roster in real time. *Comput Oper Res*, pp. 2077–2094, 2005.
- [7] Lova, A., Tormos, P., Barber, F., Ingolotti, L., Salido, M. & Abril, M., Intelligent train scheduling on a high-loaded railway network. *Lecture Notes in Computer Science, (to appear)*, 2006.
- [8] Schloegel, K., Karypis, G. & Kumar, V., Graph partitioning for high-performance scientific simulations. *Sourcebook of parallel computing*, pp. 491–541, 2003.
- [9] Karypis, G. & Kumar, V., A parallel algorithm for multilevel graph partitioning and sparse matrix ordering. *Journal of Parallel and Distributed Computing*, pp. 71–95, 1998.
- [10] Yokoo, M. & Hirayama, K., Algorithms for distributed constraint satisfaction: A review. *Autonomous Agents and Multi-Agent Systems*, **3**, pp. 185–207, 2000.
- [11] METIS, <http://www-users.cs.umn.edu/karypis/metis/index.html>.



This page intentionally left blank

Blocking time reduction for level crossings using the genetic algorithm

Y. Noguchi¹, H. Mochizuki¹, S. Takahashi¹, H. Nakamura¹,
S. Kaneko¹ & M. Sakai²

¹*Nihon University, Japan*

²*The Nippon Signal Co., Ltd., Japan*

Abstract

The blocking time of a level crossing influences traffic on a road that crosses a rail line. The blocking time of a level crossing will be especially long on a heavy traffic rail line. The purpose of this paper is to reduce the blocking time of level crossings by optimising the railroad schedule. We propose an optimal schedule in which the departure time at each station is delayed minutely from the time of the planned schedule. Since there are many trains on a heavy traffic rail line, the number of combinations of minutely adjusted departure times will be enormous. We used the genetic algorithm (GA) for searching for an optimal railroad schedule. The delay time for each train at each station is used as a gene value, and a chromosome is composed by connecting the genes for each train. The fitness value is the total blocking time of all level crossings on the model rail line. Results for the computer simulation show that our optimal railroad schedule gives a shorter total blocking time compared with the planned schedule. In addition, we report our examination of a composition of the GA calculation with a grid computing technology for the purpose of reducing the GA operation time

Keywords: level crossing, railroad schedule, genetic algorithm, grid computing.

1 Introduction

At present, on Japanese railroads, an increase of passenger transportation capacity is brought about by an increase in the number of trains and an increase in the number of cars to a train. This leads to lengthening the time that a level crossing is blocked by trains going up and down. As a result, there is a problem



that a level crossing is often kept blocked for a long time. There are some reported measures to meet this problem, such as an overhead crossing, but the cost of those measures is high [1].

In this paper we examine a way to construct a railroad schedule in view of reducing the blocking time of level crossings. Specifically, we calculated an optimal schedule that minimizes blocking time by varying the time duration for a train to be standing at a station. And because the number of combinations of the time duration is enormous, in the calculation of an optimal schedule we applied the genetic algorithm (GA) that is used for optimization of the problem of combinations.

In addition, in order to reduce the GA operation time, we examined the composition of the calculation system with a grid computing technology that is attracting attention as one of the distributed processing methods.

2 Optimization method of a railroad schedule in view of reducing blocking time of level crossings

2.1 Definition of blocking time of a level crossing

A level crossing is controlled by two controllers, one of which is installed at the point 1 km before the level crossing and the other at the point just after the level crossing, as shown in Figure 1. And the time that a train is running between the two controllers is defined as the blocking time of a level crossing.

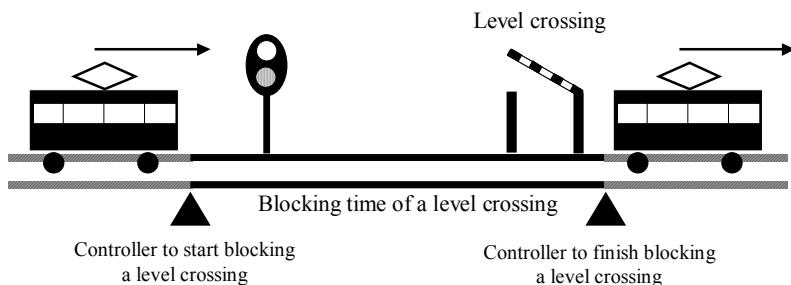
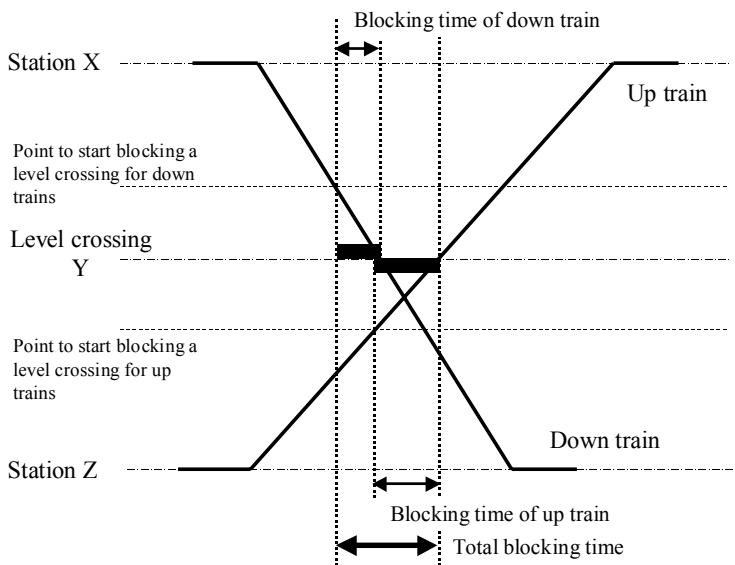


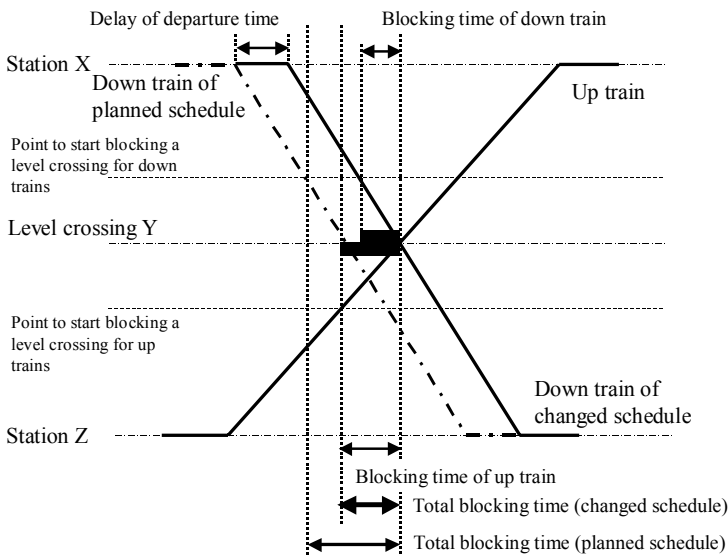
Figure 1: A block diagram of the control of a level crossing.

2.2 Outline of the proposed method and the application of GA

To calculate an optimal schedule that minimizes blocking time, we focus in this paper on a delay of the departure time at each station. As shown in Figure 2, two trains, an up train and a down train, pass through a level crossing in such a way as to maximize the time duration of the two trains being on the crossing simultaneously by delaying the departure time of the planned schedule, which results in a reduction of the blocking time of the level crossing.



(a) Planned schedule



(b) Changed schedule

Figure 2: A schematic diagram of the proposal method.



Here, for the purpose of securing the time for passengers to get on and off the train, the minimum standing time at each station was set at 20 seconds. To this we added a delay time for each train at each station in order to optimise the railroad schedule. Incidentally, because of this addition of a delay time and because of the fixed running time between stations, the arrival of a train at its terminal station will be delayed.

Now that we have shown the optimization method of a railroad schedule in view of reducing the blocking time of level crossings, we next examined the number of combinations of delay time. Because the departure time of each train at each station varies, the number of combinations in the proposed method is huge. For example, when the delay time is varied in five patterns (20[sec]+0[sec], +5[sec], +10[sec], +15[sec], and +20[sec]), and the number of up and down trains is each 20, and the number of stations is 10, then the number of combinations of departure time is 3.87×10^{279} . It is impossible to calculate all combinations in real time; therefore, we applied the genetic algorithm (GA), a powerful tool for the optimization of the problem of combinations, in our attempt to optimize the railroad schedule.

2.3 The model rail line and GA parameters

This time we used a model rail line that has 9 stations and 14 level crossings as shown in Figure 3 and Table 1. By using the delay time for each train at each station as a gene value of GA, we calculated the total blocking time of all level crossings per hour by means of the train simulator that we developed. This total blocking time was used as the fitness value.

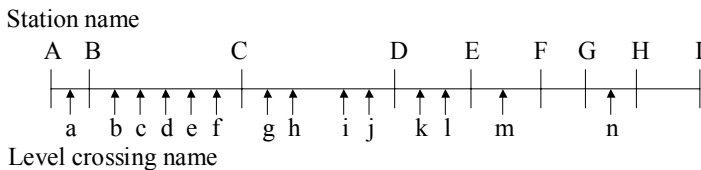


Figure 3: The model rail line.

Table 1: Distances on the model rail line.

(a) Stations		(b) Level crossings			
Name	Distance	Name	Distance	Name	Distance
A	0[km]	a	0.221[km]	j	3.643[km]
B	0.8[km]	b	0.934[km]	k	6.555[km]
C	1.7[km]	c	1.084[km]	l	6.931[km]
D	4.5[km]	d	1.232[km]	m	7.680[km]
E	7.2[km]	e	1.420[km]	n	9.937[km]
F	8.5[km]	f	1.651[km]		
G	9.7[km]	g	1.867[km]		
H	11.2[km]	h	2.386[km]		
I	13.4[km]	i	3.415[km]		



3 Verification of effectiveness of the proposed method by the use of a computer simulation

To verify the effectiveness of the proposed method shown in the previous chapter, we carried out a computer simulation this time. Specifically, after we constructed railroad schedules on the model rail line with the running interval of the trains varying from 3 to 10 minutes, we developed the train simulator using GA to calculate the blocking time per hour of a level crossing in order to search for the combination of departure times that minimizes blocking time.

When we carried out the computer simulation in which the GA parameters were set so that the number of individuals was 30 and the number of generations was 1,000, we obtained the resultant characteristic of the total blocking time of all level crossings by changing the train intervals shown in Figure 4. Figure 4 affirms that we can reduce the total blocking time by changing the combinations of departure times. Especially, it shows a significant effect of reduction of the total blocking time by 2,500 seconds when we set the train interval to be 3 minutes with up and down trains crossing each other frequently. Moreover, under this condition, the delay of arrival time at the terminal station relative to the planned schedule was about 50 seconds on the average.

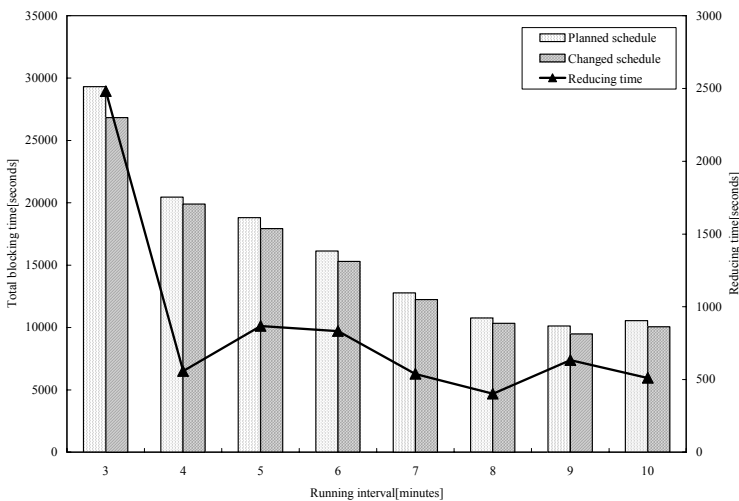


Figure 4: Characteristics of total blocking time by varying train intervals.

4 Examination of the composition of real-time calculation of optimal railroad schedule system using grid computing

4.1 Need for real-time calculation of the optimal railroad schedule

In the preceding chapter, we showed that we can reduce the total blocking time by changing the combinations of departure times. The time needed for



calculation is about 1 hour when the total number of individuals to be calculated is 30,000 (30 [individuals] $\times$ 1,000 [generations]). On the other hand, an actual railroad schedule on which trains run might be different from the planned schedule because the actual schedule is influenced by delays of departure due to passengers' movement, varying running times between stations due to weather, and so on. Based on this fact, we conclude that we must develop a way for real-time calculation of the optimal railroad schedule that can respond in real time to an actual running situation if we are to make practical use of this system. So we attempt to devise a way to speed up the present railroad schedule calculation.

4.2 Outline of the grid computing technique

To speed up the railroad schedule calculation, we used the grid computing technique, one of the distributed computing techniques. A grid computing system is composed of a master PC and some slave PCs, as shown in Figure 5. In actual processing, the master PC divides the processing that we want to be carried out into batches of a constant number of jobs, and the jobs are transferred to the slave PCs via the network. The slave PCs carry out their given jobs and return the results to the master PC. After repeating these operations, the master PC brings the processing results of the jobs together at the end. The grid computing technique offers some advantages, for example, an easy installation, low cost, and so on, so the grid computing technique is being used in various fields now.

4.3 Composition of real-time calculation system of the optimal railroad schedule

Based on the image chart that has been shown in Figure 5, we constructed a real-time calculation system of the optimal railroad schedule shown in Figure 6. In Figure 6, a GA engine plays the role of a master PC, and train simulators play the role of slave PCs. The actual processing goes as follows. A master PC that functions as the GA engine configures a constant number of departure time patterns for each station as initial genes and transfers them to slave PCs. The slave PCs that function as the train simulators calculate the total blocking time of all level crossings based on departure time patterns and return their results to the master PC. Then the master PC configures new departure time patterns by GA operation based on calculation results from the slave PCs and transfers them to slave PCs again. These operations are repeated.

4.4 Restructuring of GA engine to allow for the grid computing technique

We have shown the composition of the real-time calculation system of the optimal railroad schedule in Figure 6. Next we examined the restructuring of a GA engine to allow for grid computing. In a customary GA engine, after train simulators finished calculating the fitness values against a constant number of genes prepared in advance, the GA engine makes new genes by GA operation, and it searches for the optimal solution by repeating these operations. Thus, the



GA operation and fitness value calculation are done by turns. On the other hand, the purpose of the composition as shown in Figure 6 is to have both the master PC that functions as a GA engine and the slave PCs that function as the train simulators be active simultaneously. So it was necessary for us to restructure the GA engine for parallel processing. Specifically, we restructured the master PC to make new genes by GA operation based on the fitness values calculated by the slave PCs for part of the departure time patterns made as initial genes. By this means the slave PCs can carry out fitness value calculations for the rest of the genes while the master PC processes GA operation, and thus this composition enables us to have efficient parallel processing. Now that we have explained the restructuring of a GA engine, next we define the number of initial genes as the number of genes made by a GA engine in advance and the number of calculation genes as the number of genes used for GA operation, and we evaluated this composition thereafter.

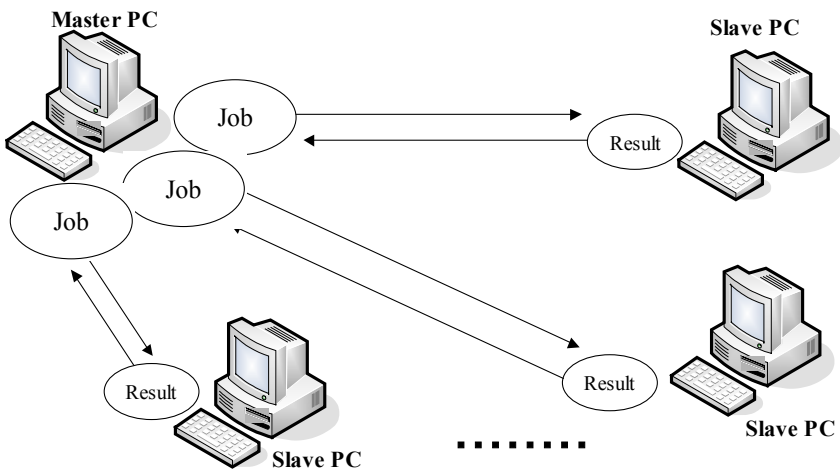


Figure 5: Image chart of grid computing.

4.5 Evaluation of a real-time calculation system of the optimal railroad schedule using a computer simulation

We evaluated the characteristics of the total blocking time when the number of calculation genes was fixed at 30, and the number of initial genes was varied in four patterns (30, 40, 59, and 60) by using a computer simulation. We show the result in Figure 7. As shown in Figure 7, no good characteristic can be obtained when the number of initial genes is 60, that is, twice the number of calculation genes, because this case is equivalent to evolution by two independent GA engines. On the other hand, good characteristics can be obtained when the number of initial genes is not 60. The characteristics when the number of initial

genes is 59 are evidently especially good compared with the characteristics of the traditional method (both the number of initial genes and the number of calculation genes are 30). In the above, we verified that this system can calculate the optimal railroad schedule most efficiently when the number of initial genes (N) is one less than twice the number of calculation genes (M). We call this the “ $N=2M-1$ method” hereafter.

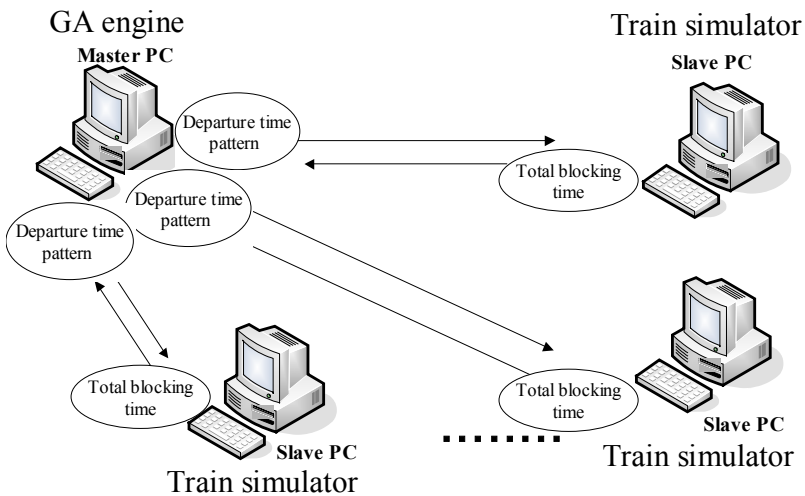


Figure 6: The composition of real-time calculation system of optimal railroad schedule.

Next, we examine the $N=2M-1$ method. We show the characteristics of the total blocking time when the number of calculation genes was varied in four patterns (15, 30, 60, and 120) in Figure 8. The numbers of initial genes for those four patterns were 29, 59, 119, and 239.

As shown in Figure 8, we verified that the characteristics of total blocking time for the number of departure time pattern calculations evolved quickly when the numbers of initial genes and calculation genes were small. And when the numbers were large, we verified that the characteristics evolved to a nearly optimal railroad schedule with a comparatively small number of calculations, even though the evolution speed was not fast. In fact, we took into consideration the characteristics as shown in Figure 8 when we mounted this system on grid computing. For example, when the number of initial genes was 120, the number of departure time calculations to optimize the railroad schedule was 15,000, which is one half of 30,000, the number needed in the traditional method. So if the number of slave PCs is 6, the optimal railroad schedule calculation time can be ideally shortened to 1/12 of the time needed in the traditional method. In sum, we have the prospects that the composition using the $N=2M-1$ method can calculate the optimal railroad schedule in a few minutes.



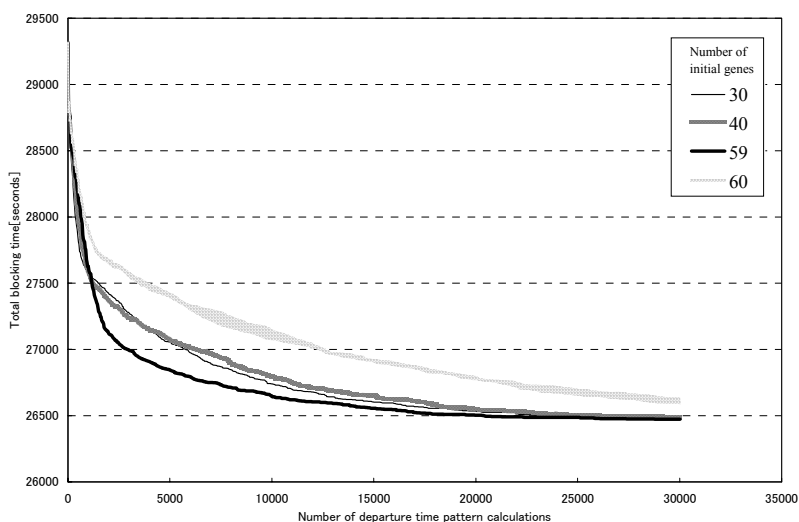


Figure 7: Characteristics of total blocking time when number of initial genes was changed.

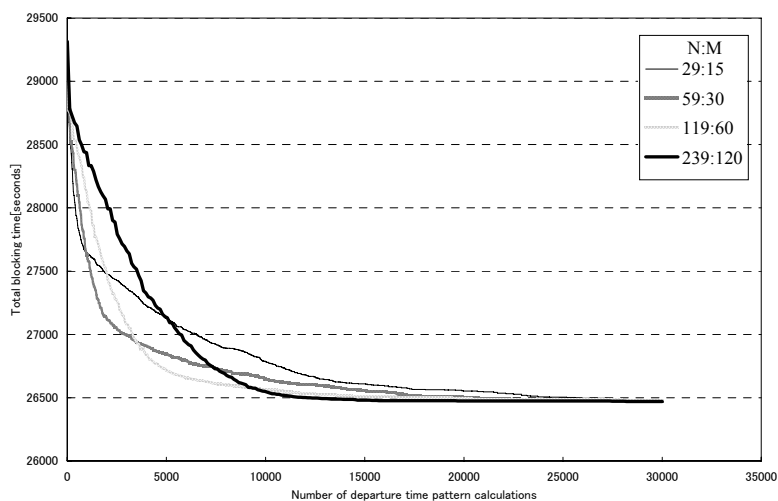


Figure 8: Characteristics of total blocking time in $N=2M-1$ method.

5 Conclusion

In this study, we examined a way to construct a railroad schedule in view of reducing blocking time of level crossings, and we proposed a method to calculate the pattern, using GA, that minimizes the total blocking time of all level



crossings by varying the time duration each train stands at each station. The result of a computer simulation verifies the effectiveness of the proposed method. In addition, we showed a composition of a real-time calculation system of the optimal railroad schedule using the grid computing technique that reduces calculation time. And as a result of restructuring the GA engine, we have the prospects that the railroad schedule should be optimized in a few minutes by the composition using the $N=2M-1$ method, in which we defined the relationship between the number of initial genes (N) and the number of calculation genes (M) as $N=2M-1$. In the future, we will implement the system that we proposed after we built a grid computing system actually using some PCs. And we hope to advance this research further by measuring the optimal railroad schedule calculation time and so on.

Reference

- [1] Atsushi SAITO, Satoru SONE and Susumu TAKANO, "A Method of Controlling a Group of Trains for Securing Opening Time at Very Heavily Trafficked Double Track Railway Crossings", IEEJ, Papers of technical meeting, TER-03-14, pp.37-42 (2003-03).



Modeling a distributed railway interlocking system with object-oriented Petri nets

X. Hei¹, H. Mochizuki¹, S. Takahashi¹, H. Nakamura¹, M. Fukuda²,
K. Iwata² & K. Sato²

¹*College of Science and Technology, NIHON University, Japan*

²*Railway Technical Research Institute, Japan*

Abstract

Great progress of distributed technology and intelligent terminals makes it possible to develop a distributed railway interlocking system (DRIS). In this paper, a modelling method of DRIS is presented by using G-nets, which are Petri nets extended with object-oriented concepts. The modelling method improves maintenance and reusability remarkably. Based on the models, the DRIS can be implemented with an object-oriented language such as C++ or JAVA.

Keywords: distributed technology, railway interlocking system, G-nets, Petri net.

1 Introduction

Computerized interlocking system has been developed for many years [1]. Conventional interlocking system, which is named electronic interlocking system, is composed of central control computer, electrical control part and mechanical part. These systems feature that cost is high and updating of terminal devices is inconvenient.

On the other hands, study and application of distributed control technology and field-control technology have become a new stage [2, 3]. Especially, with the development of ubiquitous technology [4], intelligent terminals or intelligent devices have been possible. All these progress make it possible developing distributed railway interlocking system (noted as DRIS).

In distributed interlocking system, all devices are intelligent, and they communicate with each other directly without centralized interlocking computer, as shown in figure 1. All these intelligent devices have functionality units which are shown in figure 2. The logics of DRIS also become different from centralized interlocking system.



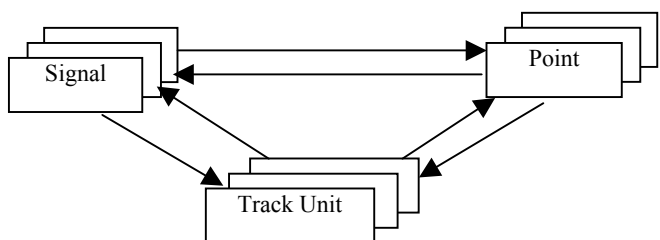


Figure 1: Distributed railway interlocking system (DRIS).

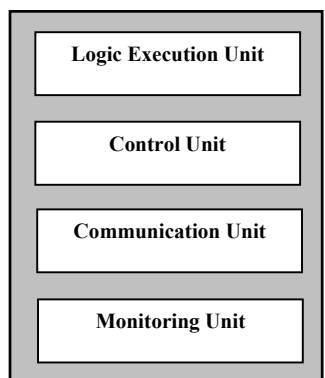


Figure 2: Components of intelligent devices.

DRIS is an autonomous decentralized and distributed real-time control system. The complexity of such system requests new system modeling techniques to support reliable, maintainable and extensible design.

At the same time, as the most important property of railway interlocking system, fail-safe has to be assured. So the correctness verification of new logic also needs a new modeling technique. Further, the dependability evaluation of DRIS is based on a proper DRIS model, which can describe the DRIS model precisely. Some efforts have been made for modeling the railway interlocking system [6, 7]. But they did not either consider DRIS or establish a uniform and hierarchical model.

In this paper, we proposed a hierarchical modeling method for DRIS with G-nets [5] which is Petri net extended with object-oriented concepts.

The reason why we model DRIS based on G-nets is that our analysis of DRIS is also based on object or class. Each device can be thought as an object, namely an independent module. They have their own functions, and they communicate with each other via some well-defined interfaces.

G-net is a modeling tool which is a Petri net extended with Object-oriented concepts, namely, the object, with which the model of DRIS can be simplified



evidently. An object packs many internal details of realization. To some extent, G-nets can be thought as a high-level Petri net.

The analysis and evaluation of G-nets model can be converted to analysis and evaluation of Petri nets, which have many available analysis methods and tools.

The paper is organized as follows: section 2 gives an overview of G-nets; modeling approach of DRIS is presented in section 3, and a simple example of station layout and its analysis are presented in section 4; finally, we conclude this paper in section 6.

2 G-nets

G-net is a Petri net based on multi-level executable specification model which incorporates the concepts of module and system structure [5, 6]. A G-nets system comprises of a number of G-nets, each of which represents an independent module. These modules communicate with each other through well-defined interfaces, that is, the methods of G-nets.

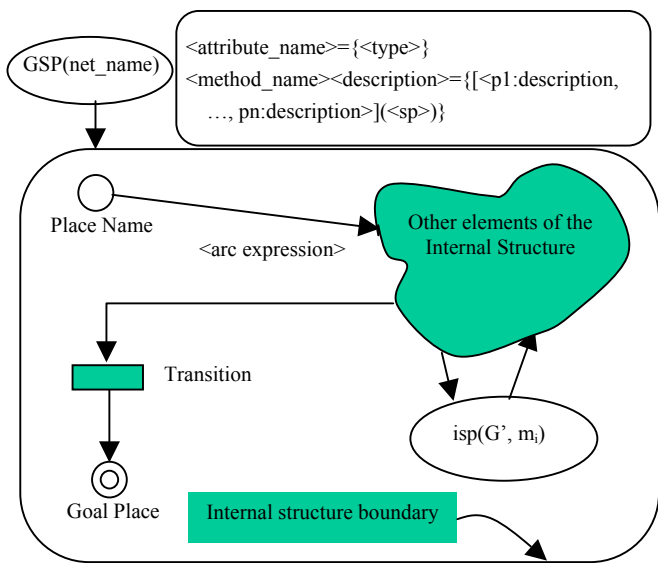


Figure 3: Notations of a G-net.

A G-net is composed of two parts: a special place called *Generic Switch Place* (GSP) and an *Internal Structure* (IS). The GSP provides the abstraction of the module, and serves as the only interface between the G-net and other modules. The IS, a modified Petri net, represents the detailed internal design and realization of the module. The notation for G-nets is shown in figure 3.

The internal structure of the net (IS) is enclosed by a round corner rectangle. The GSP is indicated by the ellipse in the left upper corner of the IS boundary. The inscription $GSP(net\_name)$ defines the name of the net. The rounded corner rectangle in the upper right corner of the IS boundary is used to identify the



methods and the attributes for the net, where: $\langle attribute\_name \rangle = \{ \langle type \rangle \}$ defines the attributes for the net, $\langle attribute\_name \rangle$ is the name of the attributes, and $\langle type \rangle$ is a type for the attribute, restricted to the set of non negative integers; $\langle method\_name \rangle$ and $\langle description \rangle$ are the name and description for the method respectively. $\langle p1:description, ..., pn:description \rangle$ is a list of arguments for the method, and $\langle sp \rangle$ is the name of the initial place for the method. An ellipse in the internal structure represents an *instantiated switch place* (*isp*), which is used to provide inter-G-Net communication, while a circle represents a normal place. The inscription $isp(G', m_i)$ indicates the invocation of net G' with method m_i . Executing the *isp* primitive implies invoking G' (by sending a token to G') based on the specified methods. This token contains the parameters needed to define the tokens for the initial marking of the invoked net. A rectangle represents a transition that may have an inscription associated with it. This inscription may be either an attribute or a firing restriction. A double circle represents the termination place or goal place. Places and transitions are connected through arcs that may carry an expression.

In the internal structure, places represent primitives, while transitions, together with arcs, represent connections or relations among the primitives. A set of special places called *Goal Place* represents the final state of the execution, and the results (if any) to be returned.

From the description above, we can see that a G-net model essentially represents a module or an object. A G-nets system supports incremental design and successive modification. We find it suitable to model DRIS.

3 Modeling interlocking system

3.1 Function modules of intelligent devices

In DRIS, there are three kinds of intelligent devices: signal, point and track unit. When a train is coming, a route will be set, related signals have to be locked in some status, track units must be in some status and related points must be locked in requested position. When all conditions of route setting are satisfied, the start signal or home signal of the route changes to permission status, which indicates the coming train can go on.

Corresponding with the components shown in figure 2, an intelligent device of DRIS is composed of 4 functionality modules, which are listed in figure 4, namely, the communication module, computing module, control module and monitoring module. Communication module allows devices to communicate with other intelligent devices or other system, such as ATS. Computing module functions as logic computing and judgment, which decides whether the logic status is correct or not, what status it should be, etc. Control module is in charge of controlling actions of devices, such as changing status, locking or unlocking. The monitoring module monitors the field device status by timely polling the device and alarms when there is abnormal status.

These modules are independent in structure and functionality, and they coordinate with each other via some interfaces.



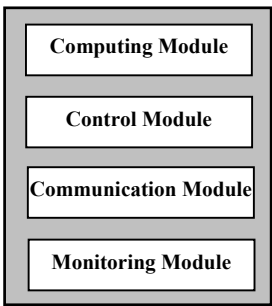


Figure 4: Modules of intelligent devices.

3.2 Modeling DRIS with G-nets

Because the communication, computing and control module of interlocking devices are similar to each other, we introduce only these modules of signal and the monitoring module of track unit. Figure 5 is G-net model of communication module of signal. There are 3 methods: *rm*(receive message), *sm*(send message) and *rr*(route request). The communication module of point and track unit has no the *rr*(route request) method. In method *sm* and *rm*, the parameters are *id* and *data*. *Id* represents network id of destination devices; and *data* represents communication data, which maybe comprises of logic status information. Parameter of method *rr* is *rid*, which indicates the id of requested route.

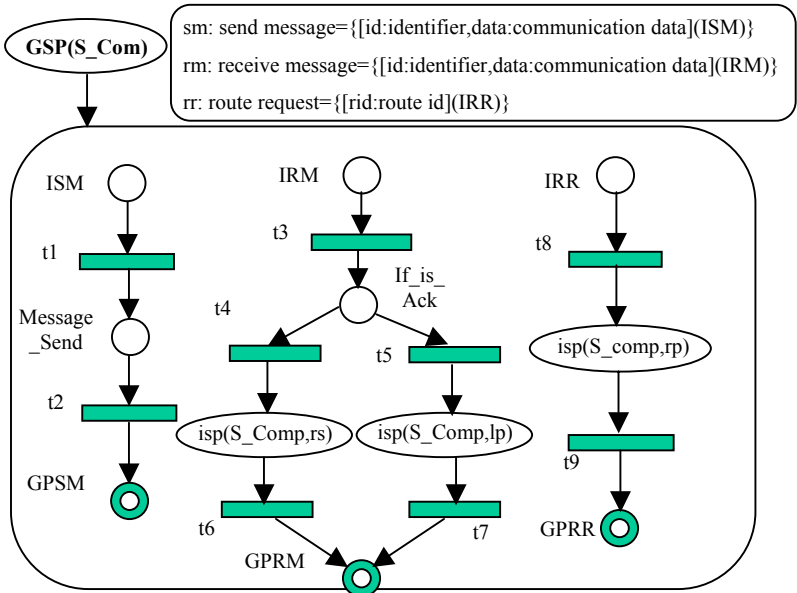


Figure 5: G-net representing communication module of signal.

In figure 5, *Message_Send* and *If_is_ack* are primitives. They can be executed without calling other methods. The return value of primitive *If_is_ack* is used to judge whether the received message is an acknowledgment from other devices, i.e. whether it is a response message or not. If it is an acknowledgment, then call *rs*(means route setting) method of module *S_Comp*, which changes the signal to permission status when all acknowledgments from route-related devices are received. Otherwise, the received message must be an order that indicates what status of signal should be. In this case, method *lp*(means logic process) of module *S_Comp* is called.

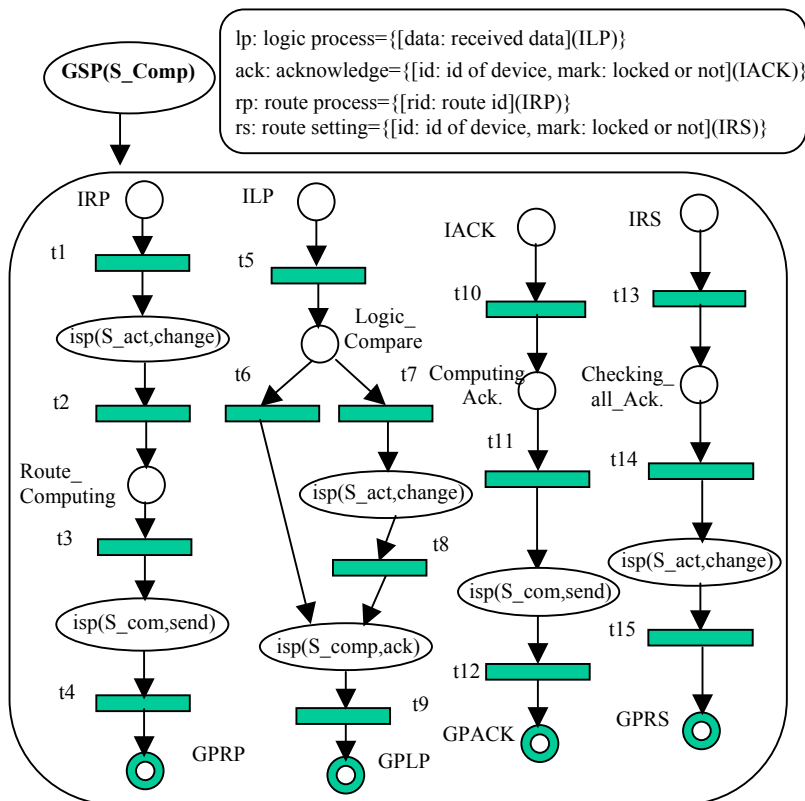


Figure 6: G-net representing computing module of signal.

Figure 6 is G-net model of computing module. There are 4 methods in module *S_Comp*. When signal receives route request via method *rr* of module *S_Com*, method *rp* changes its status to *route setting*, and computes the status information of other route-related devices, then send the result to these devices by *isp(S_com,send)*, which call *send* method of module *S_Comp*. Method *lp* executes logic processing. There are 2 circumstances that are considered when receiving message, one is the logic status coming from network is the same as current logic status (transition t6 is fired); the other one is different. In the second



case, it is necessary to change logic status and lock it. Finally, devices acknowledge a message by invoking method *ack* of module *S_Comp* to the signal that starts *route setting*. The *rs* method starts when an acknowledgment is received from other devices. When all responses are received and they satisfy the requirements (primitive *Checking_all_Ack* checks the acknowledgment message), signal change itself to *permission* (display green).

The computing module of point and track unit is different from that of signal. They have no method *rp*(route process) and *rs*(route setting).

Figure 7 is G-net model of the control module which is named *S_Act*. In module *S_Act*, there are 3 methods: *change*, *lock* and *unlock*. Method *change* changes status of device; *lock* locks the logic status of device, such as locking point on reverse or normal; *unlock* carries out unlocking action of device.

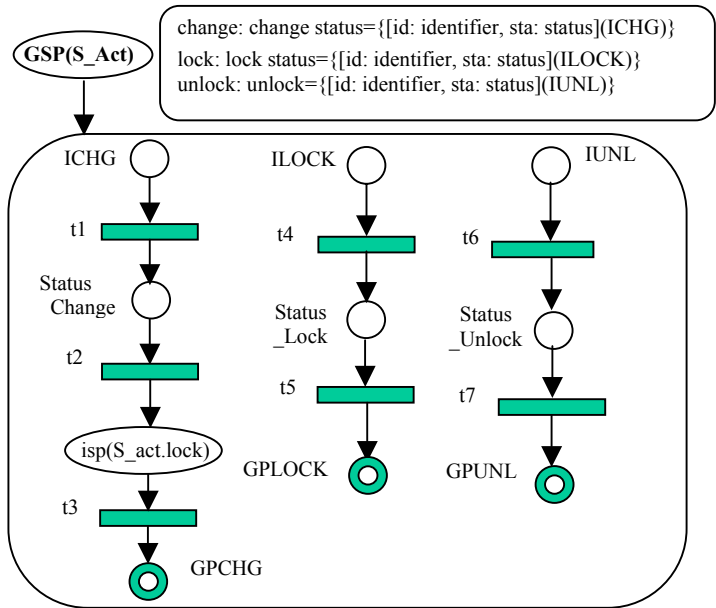


Figure 7: G-net representing control module of signal.

Figure 8 shows the monitor module of track unit. There are 3 methods in the monitor module, namely, *CheckTrain*, *NotifySignal* and *CheckPoint* methods. *CheckTrain* method checks the status of train(whether the train is on the track), *NotifySignal* method sends train status to signal, and *CheckPoint* inspects the status and position of point which is on some tracks.

These modules show the inter-module interface as well as internal realization of the module. By introducing G-nets model with data domain, the system can be represented evidently. The models can be used to develop DRIS system with any object-oriented programming language, such as C++ or JAVA, etc.



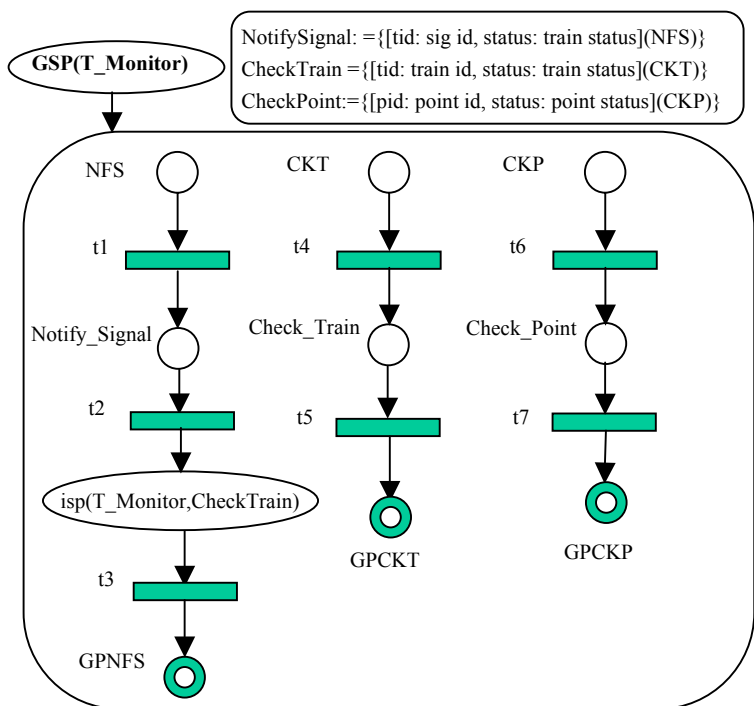


Figure 8: G-net representing monitor module of track unit.

3.3 The model architecture of DRIS

The primitives, which are shown as a circle in the G-nets model of DRIS, can be analyzed further. Thus a hierarchical Petri nets model with several levels can be established until every possible state is represented.

With G-nets, the DRIS can be analyzed at any level, which makes the DRIS G-net model to become a uniform model.

4 Analysis of a DRIS example

Figure 9 is an example of DRIS, which is used to describe the modeling approach. There are 8 routes, 4 home signals, 4 start signals, 2 points and 2 platforms. Signal has 3 statuses: permission, refusal, and route setting. Table 1 is interlocking information table which is related with route X→1. From figure 9, we find that signal S1 is the home signal of route X→1, and when X→1 is requested, signal S3 must be locked in refusal status.

As a simple illustration, figure 10 shows the invocation process of signal S1 and signal S3 when route X→1 is requested. When X→1 is requested, the route request is sent to signal S1. S1 changes its status to route setting, then searches route related devices and computes what logic status the devices should be. When the process is completed, signal S1 send the computing result to network.



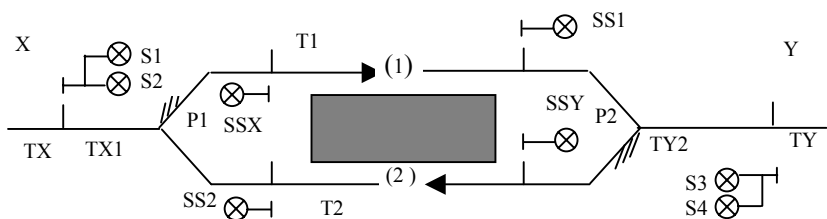


Figure 9: Layout of interlocking example system.

Table 1: Interlocking information related with route X→1 in figure 9.

Route	Signal display state	Point state	Related track unit
X-1	S1-G, S3-R, SS2-R	P1-N, P2-R	TX, TX1, T1

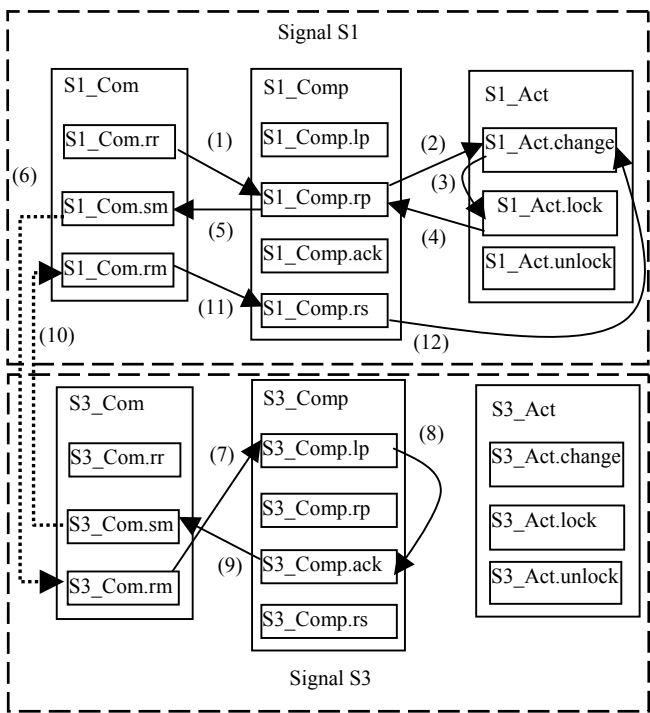


Figure 10: Invocation flow of signal S1 and S3 when X-1 is requested.



These steps are shown from step (1) to step (6). When signal S3 receives the message from network, it analyzes the message and checks whether it is the same as the current status or not. In figure 10, we assume they are the same, i.e., signal S3 is refusal status, so acknowledgment message is sent back to S1 directly. Step (7)→(10) shows this process.

When signal S1 receives the acknowledgments from all route-related devices and if they satisfy the requirements, i.e., all conditions for signal S1 changing to permission are achieved, signal S1 changes itself to permission status (display green). Step (11) and step (12) describe this process.

Step (6) and step (10) represent the communication between signal S1 and S3, which is accomplished by using a network.

Processes of other devices such as signal SS2, point P1 and P2 are similar to process of signal S3.

5 Conclusion and discussion

The proposed approach can improve greatly the reusability of function module when designing and realizing the system, and can simplify the complexity of real-time control system when analyzing these systems.

In future research, the reliability of G-nets DRIS models will be analyzed with Petri net tools and methods.

References

- [1] K. Akita, T. Watanabe, H. Nakamura, I. Okumura: "Computerized Interlocking System for Railway Signaling Control; SMILE", *IEEE Trans., Ind.*, 1A-21, May 1985.
- [2] Heck, B. Wills, L. and Vachtevanos G. "Software Technology for Implementing Reusable, Distributed Control. Systems", *IEEE Control Systems Magazine*, February 2003.
- [3] C. Engelmann, S. L. Scott and G. A. Geist. "High Availability through Distributed Control". *Proceedings of High Availability and Performance Computing Workshop (HAPCW)*, Santa Fe, NM, USA, October 2004.
- [4] M. Weiser and J. Brown, "Designing Calm Technology," *PowerGrid J.*, vol. 1, 1996.
- [5] Y. Deng, S. Chang, A. Perkusich and J. de Figueiredo, "Integrating Software Engineering Methods and Petri Nets for the Specification and Analysis of Complex Information Systems". *Proc. Of The 14th International Conference on Application and Theory of Petri Nets*, Chicago, June 21-25, 1993, pages 206-223.
- [6] J. de Figueiredo and A. Perkusich, "Distributed Control of Track-Vehicle System with Fault-Tolerant Characteristics: a Petri Net Based Approach", *Proc. of the IEEE international conference on Systems, Man, and Cybernetics*, paginas,377-382,Vancouver, Canada, 1995
- [7] V. Hartonas-Garmhausen, "Verification of a safety-critical railway interlocking system with real-time constraints", *Proc. of 28th Annual Int. Symposium on Fault-Tolerant Computing*, p.458, June 23-25, 1998.



Using UML diagrams for system safety and security environment analysis

F. M. Rachel & P. S. Cugnasca

Safety Analysis Group, Department of Computer Engineering and Digital Systems, Polytechnic School, University of São Paulo, Brazil

Abstract

In this paper, the use of UML (Unified Modelling Language) diagrams as software tools for system safety and security environment analysis is proposed and evaluated.

The UML diagrams are used to plan and build systems based on the Object-Oriented approach. As these diagrams allow many system aspect viewpoints, they also allow a deep analysis and understanding of the system architecture and implementation details, as well as system functioning and operational features.

When a system safety/security environment analysis is accomplished, many aspects of system operation, functioning, data flow, data types, architecture and implementation details must be well known, understood and modelled in order to determine possible weak points for the system safety (or security, or both, depending on the system application). The various UML diagrams supply all the information needed for a safety/security system analysis and many aspects of the UML methodology can be applied for the same purpose.

Finally, a case study for an Object-Oriented ATO (Automatic Train Operation) control system proposed for use on a subway system is conducted in order to analyse the safety environment and to identify possible risks and danger situations to the system operation.

This control system proposed was presented and discussed in COMPRAIL 2004, in the paper called “Object-Oriented Approach for Automatic Train Operation Control Systems”; now this paper presents a complement of that study, using the drawn diagrams to make an analysis of the system safety environment.

Keywords: automatic train operation, control systems, train control, object-oriented project, object-oriented analysis, UML diagrams, safety-security analysis.



1 Introduction

In 2003, a study was started in the Safety Analysis Group at the Polytechnic School of São Paulo University to make a viability analysis of the use of predictive fuzzy logic in an automatic train control for subway systems. In order to do this viability analysis, there were many aspects to be considered.

One aspect of this study was the use of the Object-Oriented (OO) approach, presented in Comprail 2004, in the paper called “Object Oriented Approach for Automatic Train Operation Control Systems”. The main reason for choosing OO techniques was the viability analysis of using the OO approach for safety critical applications, such as subway transportation control systems [1].

In that study, a proposal for an object-oriented project was developed for an Automatic Train Operation (ATO) system. UML diagrams were used to specify all project dimensions – static, dynamic and method dimensions; all project phases were presented and detailed in the UML diagram forms [1].

Another study aspect considered was the use of predictive fuzzy logic itself. As this tool is an Artificial Intelligence (AI) tool, some considerations had to be taken into account in order to make the project comply with the IEC 1508 standard (Functional safety of electrical/electronics/programmable electronic safety-related systems), because subway transportation is a safety-critical system [2].

Some subway systems, such as those used in the Japanese city of Sendai and in the Brazilian city of São Paulo, were chosen as models for this proposed system design. The Sendai subway uses fuzzy logic and São Paulo subway uses track circuits to control train movement in a fast and safe way.

Another study aspect considered was the safety aspect. As a subway control system, it is fundamental that a risk analysis is conducted on this kind of system. Although the system has not yet been built and implemented, it is possible to make a preliminary risk analysis in order to detect weak points to be improved in the system final design [3].

Later in the study, some improvements took place based on the preliminary project; the study presented in this paper aims to show these improvements, incorporated to the project, and to point whether is possible to make a safety/security environment analysis based on UML diagrams that describe the system design and functioning by means of its project implementation characteristics.

The improvements and the adaptations needed to be updated and to make the system comply with the IEC standard lead to a system architecture rebuilding, showing that it is possible to make an ATO controller using AI tools, maintaining the system safety operating margins. However, some changes in the system architecture will take place in order to accomplish this task [2].

At the moment, this project is still not implemented in any subway system and there are some points to be analysed in future researches in this area, as the incorporation of other ATO functions to the project, such as program stop, doors opening/closing and data communication between the train and the station.



2 The updated object-oriented system proposed

In this section, the updates to the original project presented in Comprail 2004 will be shown. In that study, one possible solution was presented, among many others, and since its beginning, in the object point of view based on the UML, a well-defined and standardized methodology developed for object-oriented projects creation was used [1]. The UML uses several diagrams for all project dimensions: static, dynamic and method. For safety applications, one more dimension is added [4]. Figure 1 shows the safety/security aspect added to the other dimensions and the respective diagrams.

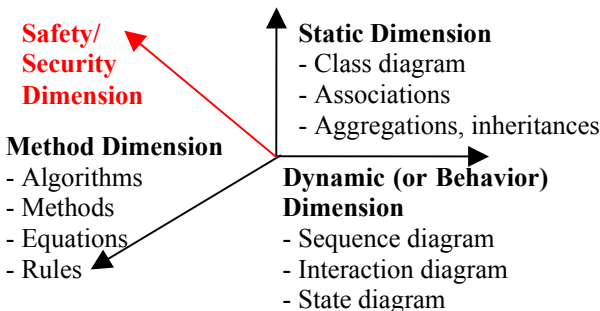


Figure 1: Safety/security dimension is added to static, dynamic and method project dimensions.

The system project and construction process have many phases, each focusing on a system aspect. The updated waterfall model including the improvements phase is shown in figure 2 [5].

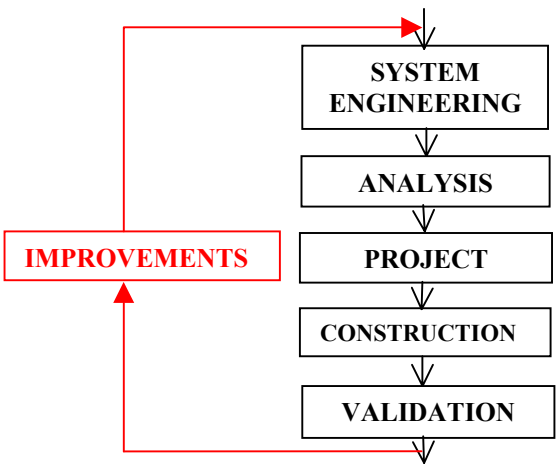


Figure 2: Waterfall model including improvements phase.



2.1 System engineering

When improvements are included in the original project, the project cycle reinitiates and, again, understanding the problem is essential to the project success. However, it is not necessary to begin from scratch, but to understand how the improvements affect the original project. In the original project, the system to be controlled was the Automatic Train Control (ATC), including two main functions - Automatic Train Operation (ATO) and Automatic Train Protection (ATP). This was not modified and figure 3 is only a reference for understanding the problem [1].

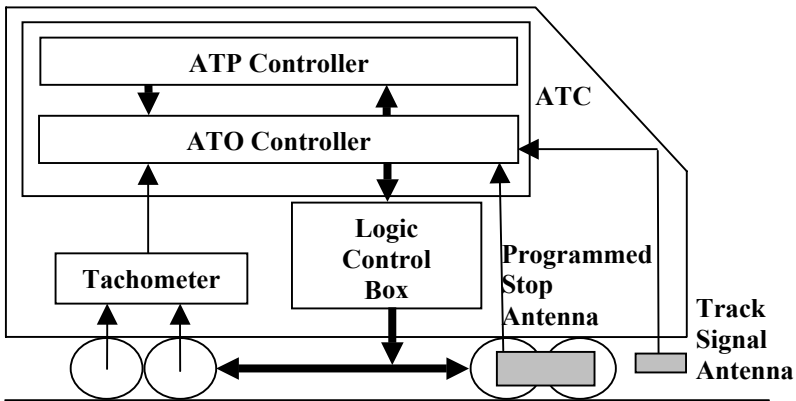


Figure 3: Schematic diagram of the ATC system [1].

The improvements to the original projects include an operational need for control system performance, according to certain operational situations. For example, the operational team has different needs for performance, according to operational demand on users. At peak times, the trains must run with shorter headways (time intervals between trains) and on valley demands the trains can run with larger headways, seeking energy consumption saving. This new problem was solved including a new entry called *Time or Performance Level* and it is related to the time interval used by the train for going from one station to another.

The updated IDEF (Integration Definition for Function) diagrams for this new configuration are shown in figures 4 and 5. As presented in Comprail 2004, they have a pattern graphical notation to represent the information flow and the processes used, in a hierarchical top-down architecture, in which processes can be expanded and detailed in hierarchical diagrams. The main process is represented in the IDEF0 top diagram called IDEF0 level 0 or context diagrams [6].

Again, the system engineering phase must detail the additional time and costs needed to implement the new solution [4].



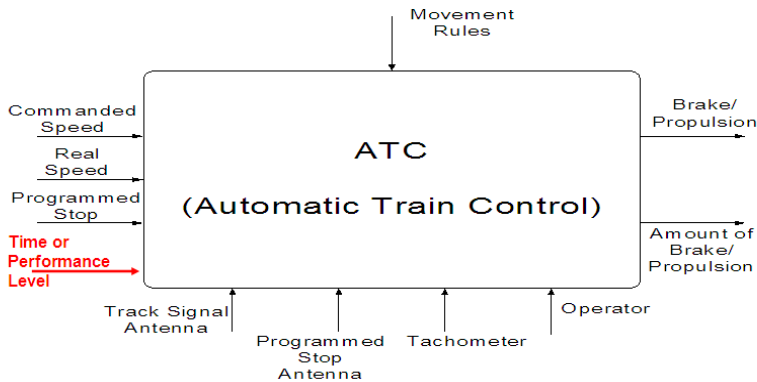


Figure 4: IDEF Level 0 (Context Diagram) including the new entry.

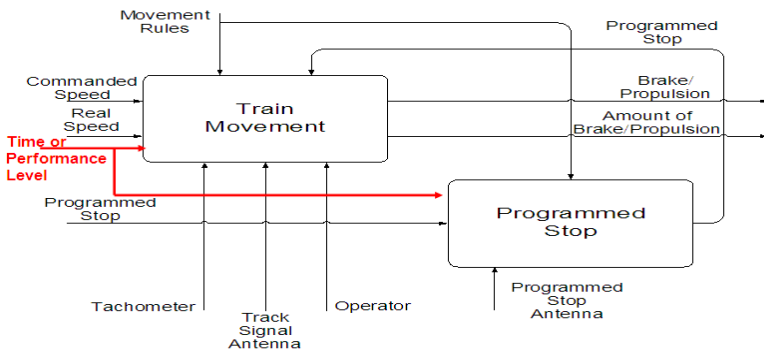


Figure 5: IDEF Level 1 including the new entry.

2.2 Analysis

At this point, besides the new solution implementation, the main objective of this paper is presented, that is, to verify whether the UML diagrams can be used to analyze the safety/security application environment. As will be shown, the answer to this question is *yes*. The static diagrams can be checked about wrong entries, out-of-range values or format errors, while dynamic diagrams can point the error consequences or predict how error scenarios can influence system behaviour. Unexpected answers shall be predicted and the system behaviour to these answers must be predicted and mapped in the diagrams.

Figure 6 shows the class diagram, including a new class (*Performance*) for new entry representation, while figure 7 shows the state diagram including an



unsafe state and figure 8 shows an example for sequence diagrams including the unsafe conditions scenarios not predicted in the previous project. Again, the analysis phase must produce a system prototype, replaced by a function simulation, in this case due to lack of screens or user interfaces [1, 7]

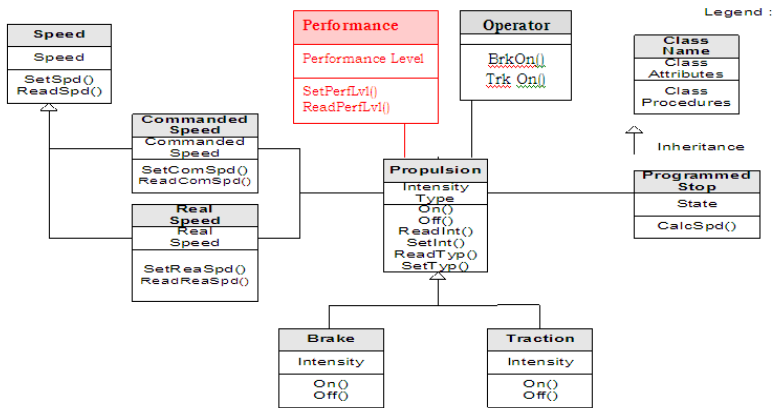


Figure 6: Class diagram including the new entry (*Performance*).

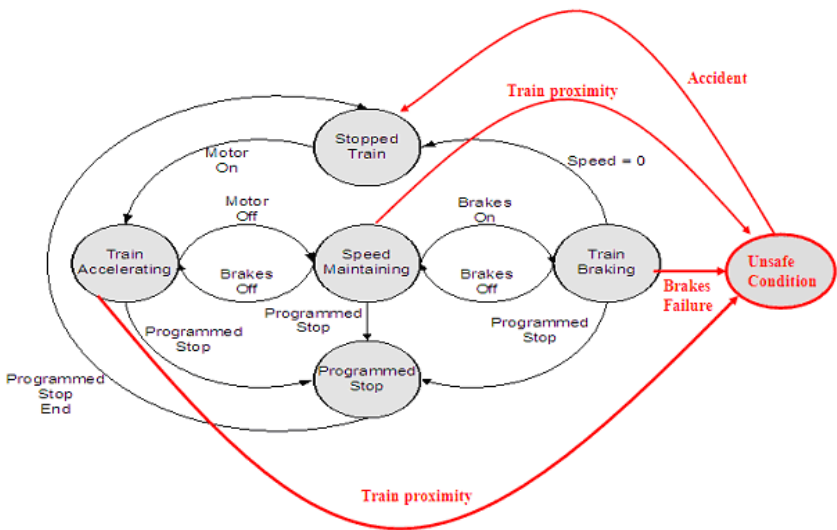
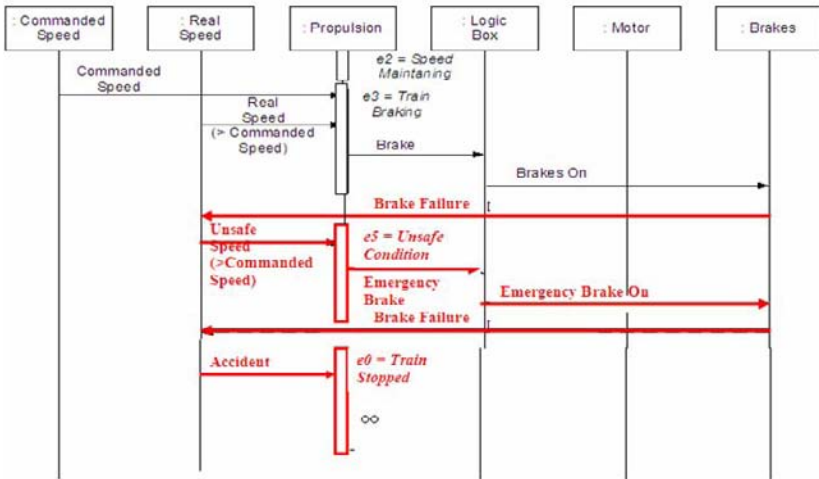


Figure 7: State diagram, including an unsafe state.



Scenario 5 – Unsafe Scenario – Brake Failure

Figure 8: Example of a sequence diagram including an unsafe scenario.

2.3 Project, construction and validation

Again, after the whole system modelling, the next step is to put the plans into practice, implementing all the functions by means of hardware and/or software [4].

The component software diagram includes the new entry class that can be either handled as a separate component or grouped to operator class and compose a new component called external interface. In the project phase, all the modules and components must be technically specified and dimensioned [4, 7].

The following phases are construction and validation. In the construction phase, the software modules are coded in a specified programming language; in the validation phase, the system is tested and put into operation.

The use of UML diagrams and techniques made the improvements implementation much easier and consistent with the original project. This was possible due to object-oriented approach characteristics that are used in UML project techniques.

Is important to stress that, even if the UML diagrams can be used to identify potential risk scenarios and to map the dangerous and unsafe consequences to the system, they do not substitute the risk analysis that must be conducted on the real system implemented, because this analysis must include real situations in real system operational environment.

Therefore, if the UML diagrams used for system safety/security environment analysis do not substitute the traditional risk analysis method, why should they be conducted? The answer is that conducting UML risk analysis will increase the system understanding in terms of safety/security and this will avoid



surprises (generally unpleasant ones) concerning system operation related to safety aspects.

3 Conclusions

Along the study development, an operational need was detected: that was the system performance control, in order to adequate system performance to the user system demand. This need required a new entry, called performance level and represented the implementation of one adjustment (or control) in the overall system performance.

The choice for the object-oriented approach for developing the system project brought the main advantage to improvements implementation: the ease of project rebuilding. This new entry implementation task was enormously facilitated due to the design approach chosen. This approach feature is very useful for new improvement implementations. However, the relation between cost and benefits provided for new solution implementation must be taken into account in order to determine whether this implementation task is worthwhile [4, 7].

The risk analysis for the project developed led to some modifications in the system architecture initially proposed. For the project to comply with the IEC 1508 standard, which aims at functional safety of electrical, electronic and programmable electronic safety-related systems, the AI tools could not be used to control critical systems related to safety such as subway transportation systems. On the other hand, the use of predictive fuzzy logic showed to be a suitable solution for control systems applications [2, 3].

The solution was encountered by separating the functional tasks from the safety monitoring tasks and implementing these tasks in separate units, one for ATP functions and the other for ATO functions. The ATP unit must be in parallel with the ATO unit and its function must be restricted to the monitoring of ATO inputs and outputs, acting with priority over ATO decisions in case these inputs and outputs lead the system to an unsafe condition.

With this separation, the ATP unit must be implemented using the traditional safety redundancy techniques such as hardware and software redundancy, modules with voting system and multi-version programs. In the latter case, the ATO functions could be used as one version, whereas the ATP could make another version for the program. One voting mechanism could produce a final action.

The ATO unit, therefore, once monitored by the ATP unit, can execute the functional tasks of train control using any AI tool, without disagreeing the rules provided in the IEC 1508 standard [2].

Another conclusion is that, even if UML diagrams can be used efficiently for safety/security system environment analysis, they do not substitute the traditional risk analysis tools, fundamental for critical safety systems such as subway transportation systems and must be done in addition to these traditional tools. The UML analysis will support the traditional system risk analysis and will increase the system safety environment comprehension, avoiding unexpected systems behaviours related to safety aspects [3].



References

- [1] Rachel, F.M., Cugnasca, P.S., *Object-oriented approach for automatic train operation control systems*, Computers on Railways IX Proceedings, Wit Press, 2004.
- [2] IEC International Eletrotechnical Comission, *Functional safety of electrical, electronic and programmable electronic safety-related systems*, IEC 1508, 1997.
- [3] Almeida Jr., J.R., Camargo Jr., J.B. and Cugnasca, P.S., *Risk Analysis of Railway Control Systems*, 8th International Heavy Haul Conference Proceedings, International Heavy Haul Association Inc., 2005.
- [4] Pressman, R.S., *Engenharia de Software, A Análise Estruturada e Suas Extensões, Análise Orientada a Objeto e Modelagem de Dados and Projeto de Tempo Real*, Makron Books, 1995.
- [5] Rumbaugh, J. et al., *Modelagem e Projeto Baseado em Objetos*, Ed. Campus, 1994.
- [6] FIPS – Federal Information Processing Standards Publication (FIPS Std 183), *Integration Definition for Function Modelling (IDEF0)*, 1993.
- [7] De Champeaux, D., Lea, D. & Faure, P., *Object-Oriented System Development, Object Dynamics and Object Interaction*, Addison Wesley, 1993.



This page intentionally left blank

System-independent and quality tested availability of railway data across country and system borders by the model driven approach

H. R. Gnägi¹ & N. Stahel²

¹*ETH Zurich, Institute of Geodesy and Photogrammetry IGP, Switzerland*

²*University of Zurich, Institute of Geography, Switzerland*

Abstract

The model driven approach attacks tasks around (geo-) data services by first exactly describing the data to be treated on the system- and format-independent conceptual level. Various services can effectively be supported by the precise conceptual model. We will show how e.g. the automatic link to a standard transfer format allows system-independent railway data checking and the successful integration of differently structured track data. A conceptual model for the railway data transferable by the XML-based transfer format railML© allows the comparison with the automatically generated XML-based model driven format. Different possibilities are presented to combine the two approaches for integrating railway data across system and country borders and providing a sustainable interoperability.

Keywords: model driven approach, data integration, interoperability, railway data, transfer format, semantic transformation, railML©, UML, INTERLIS.

1 Introduction

Our application of the model driven approach to railway data started in summer 2003. The question was, if and how very differently structured administrative and commercial data from various systems could commonly be made available without changing the original systems. Especially needed was the link of these data with the track geometry. Hofmann [4] arrived to link CERES and DfA data and to provide a common visualisation. In the LINKOST project of Swiss Federal Railways SBB [2] and the GeoRail project of UIC [1] N. Stahel and H.R. Gnägi [7] provided a demonstrator implementation extending Hofmann's data



hub. The possibility to inspect the data behind the graphical representation was added to the data hub in a client-server-environment. In the GeoRail project we have as well been confronted with the task to integrate German DB track data into the Swiss SBB system and vice versa given different coordinate systems, data structures and exchange formats on each side. Again the restriction had to be respected not to change any of the two given systems. The solution of this problem mainly on a system- and format-independent level will be treated later in this paper (section 4).

Data transfer using the model driven approach starts always with an analysis and a system- and format-independent exact description of the structure of the data to be transferred. This description is called conceptual data model or conceptual schema. The transfer format (description) is then automatically derived from the conceptual schema. We have therefore been interested to investigate the data structure behind railML© an exchange format for railway data described by XML-schema. We provided a conceptual schema of this data structure and automatically derived an XML-transfer-format from it. We will report the results of the comparison of railML© with the automatically derived format (section 3).

Section 2 gives a short introduction to the model driven approach (MDA) and its actual applications to railway problems, section 5 provides an overview of ongoing work and section 6 conclusions.

2 The model driven approach (MDA)

The model driven approach (MDA) together with semantic transformation applied to the exchange of differently structured data with the same content can shortly be described by the following four phases:

- (A) Before any data are reformatted or exchanged or treated by any other service, the structure of these data is exactly described on the system- and format-independent conceptual level. This description by a conceptual schema language is called data model or conceptual schema. A widely used graphical conceptual schema language (CSL) is the unified modelling language (UML).
- (B) From this conceptual schema, once tested for syntactical correctness by a compiler, is automatically calculated the description of the corresponding standard transfer format according to fixed rules (e.g. the description of an XML transfer format like railML(C) by the XML-schema language).
- (C) The data of the systems concerned are reformatted from the proprietary format of their systems to the standard format corresponding to their conceptual schema, done by a piece of software called 1:1-processor. This is a rather easy task, because the two formats correspond to the same data structure: The original format was the basis for the conceptual schema and from this the standard format has automatically been produced.
- (D) Given the conceptual schema and the data for the different systems, there exists a set of system-independent tools e.g. to automatically check the



quality of the data (a checker can compare the data with the corresponding conceptual schema), for data transfer even between differently structured systems (by conversion systems allowing the mapping of data models and by existing 1:1 processors for different standard data formats), or – as side effect – for a well documented data save (data model plus data in corresponding standard format)

Phases (A) and (B) are part of the core of the MDA corresponding to the classical four shell modelling procedure of the data base design. From the description of the reality selection by natural language, taxonomies and thesauri (shell 1) is developed the conceptual schema (or conceptual model, shell 2, this is our phase A). The conceptual schema has to be transformed to the possibilities of an actual data base according to its logical elements, with the logical schema as result (or logical model, shell 3). In the case of data base or GIS generation, the physical level is normally no more described by a physical schema (or physical model, shell 4) but automatically implemented by the data base software given the logical schema. Quite different is the case of data exchange: There no logical schema is needed but a physical or format schema (or format model, shell 4, our phase B) describing the transfer format. This format schema should automatically be derivable from the conceptual schema.

Phase (D) allows – among others - restructuring data of the same content by mapping different conceptual schemas. We have used this option to integrate railway data. This so called semantic transformation needs not only the conceptual schemas of the data structures concerned and the mapping of them, but in addition the data in a standards format corresponding to the conceptual schemas.

Phase (C) builds the bridge between the given mostly proprietary format and the needed standard format for the starting data, and vice versa for the target data.

We are using the following tools and transfer formats in the four phases:

- (A) The conceptual schema languages UML (graphical, for overviews) and INTERLIS 2 (textual, object-oriented, for detailing data types, constraints etc.), the UML-INTERLIS editor and the INTERLIS compiler [5]
- (B) In the railML© section 3 we use the format INTERLIS-XML. In the transformation and transfer section 4 we use ITF. The following format descriptions are automatically available from the INTERLIS compiler:

Format	Format Description Language
INTERLIS Transfer Format ITF	fmt-File (proprietary)
INTERLIS XML	XML-Schema
GML (Geography Markup Language)	XML-Schema

- (C) We had to write the 1:1 processors ourselves, because the very special railway formats we had to deal with are not yet treated by standard software.
- (D) For semantic transformation we used the INTERLIS conversion system ICS.



We will apply in section 3 the phases (A) and (B) to compare the manually developed XML-schema description of railML© with the XML-schema description automatically derived from a corresponding conceptual schema. In section 4 the whole scale of phases (A) to (D) will be used to exchange railway data.

3 The railML© transfer format automatically generated from a conceptual model

The initiative railML.org started in 2001 because of difficulties to link different railway software. railML© (railway markup language) is a common evolutionary project of railways, software and consulting enterprises as well as international research institutions. The Fraunhofer institute for traffic and infrastructure systems (IVI) in Dresden (Germany) is administrating and coordinating the partnership [6]. railML© is an XML-based exchange format, whose description (the physical or format schema, see section 2) is given in the description language XML-schema.

```
<?xml version="1.0" encoding="UTF-8"?>
.....
<!--railML: element definitions: -->
.....
<!--Infrastructure: track data: -->
<xsd:complexType name="trackType">
  <xsd:sequence>
    .....
    <xsd:element name="trackTopology" type="trackTopologyType">
      <xsd:element>
        .....
      </xsd:element>
    </xsd:sequence>
    <xsd:attribute name="trackID" type="xsd:string" use="required">
      </xsd:attribute>
    <xsd:attribute name="type" use="optional">
      <xsd:simpleType>
        <xsd:restriction base="xsd:string">
          <xsd:enumeration value="mainTrack"/>
          <xsd:enumeration value="secondaryTrack"/>
          <xsd:enumeration value="connectingTrack"/>
          <xsd:enumeration value="sidingTrack"/>
          <xsd:enumeration value="stationTrack"/>
        </xsd:restriction>
      </xsd:simpleType>
    </xsd:attribute>
    <xsd:attribute name="mainDir" type="dirValidityType" use="optional">
      </xsd:attribute>
    <xsd:attribute name="trackName" type="xsd:string" use="optional">
      </xsd:attribute>
    </xsd:complexType>
  .....
</xsd:schema>
```

Figure 1: XML-schema of railML©.

From the point of view of the MDA this development is located in phase (B) and it would be interesting to have a conceptual description of the data “behind”



the transfer format, as provided by phase (A). Therefore we started with the XML-schema description of the format (see figure 1) together with an example (corresponding part of XML-data in figure 2).

Analysing the XML-schema description of railML© and taking into account the example data, we have been able to get a graphical conceptual schema as UML-diagram (see figure 3).

```
<?xml version="1.0" encoding="UTF-8"?>
<railml xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance
  xmlns=www.railml.org/schema/infrastructure/v100
  xsi:schemaLocation="http://www.railml.org/schema/infrastructure
/v100 infrastructure_v100.xsd">
  <infrastructure version="1.00">
    <lines>
      .....
      <line lineID="DA" lineName="D-C-B-A">
        <infraAttrGroupID>1</infraAttrGroupID>
        <lineDescr>Line D-C-B-A: Mainline</lineDescr>
        <tracks>
          .....
          <track trackID="D-1" type="stationTrack"
            trackName="Station D; Stationtrack 1" mainDir="both">
            <trackDescr>Station D; Stationtrack 1</trackDescr>
            <trackTopology>
              <trackBegin>
                <bufferStop elemID="StartTrackD-1" pos="0.000"
                  absPos="84.250"/>
              </trackBegin>
              .....
              <crossSections>
                <crossSection pos="0.125" absPos="84.125" ocpIDRef="D"
                  ocpTrackID="1"/>
              </crossSections>
            </trackTopology>
          </track>
          .....
        </tracks>
      </line>
    </lines>
  </infrastructure>
</railml>
```

Figure 2: RailML© data corresponding to the XML-schema of figure 1.

A main problem of this first part of phase (A) was to get an object-oriented overview of the heavily tagged and deeply nested XML-schema text. In the second part of phase (A) the INTERLIS-UML-editor provided automatically from the UML-diagram a skeleton of the corresponding textual conceptual schema in INTERLIS (2) CSL. We completed this draft INTERLIS schema by attribute types, especially by geometric ones, by value constraints etc. From this precise conceptual schema the INTERLIS compiler is able to automatically derive the description of a corresponding transfer format. In figure 5 is shown the part of the XML-schema description of the INTERLIS-XML format corresponding to the XML-schema of the railML© format of figure 1.



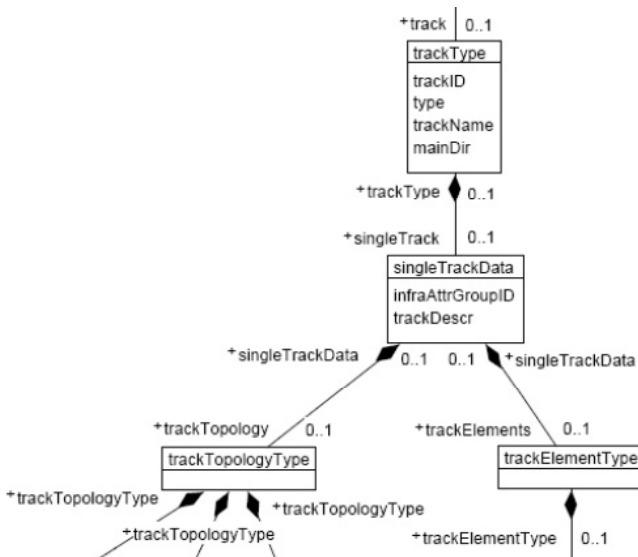


Figure 3: Part of the graphical conceptual schema for railML© as UML-diagram.

```

INTERLIS 2.2;
MODEL railML =
!! Model corresponding to railML/infrastructure
DOMAIN
  dirValidityType = (none, up, down, both, unknown);
  ....
TOPIC infrastructure =
  ....
  STRUCTURE singleTrackData =
    ....
    trackTopology: trackTopologyType;
    ....
  END singleTrackData;
  STRUCTURE trackType =
    trackID: MANDATORY TEXT*20;    !!Unique track ID
    type: (mainTrack,secondaryTrack,connectingTrack
           ,sidingTrack,stationTrack);
    trackName: TEXT*20;
    mainDir: dirValidityType;      !!Main direction of mileage
    singleTrack: singleTrackData;
  END trackType;
  ....
END infrastructure;
END railML.

```

Figure 4: Part of the textual conceptual schema for railML© in INTERLIS (2) CSL.

Comparing the XML-schema automatically derived from the conceptual model and the original XML-schema, which was the basis for the definition of the conceptual model we can remark:

- The main structure of the two XML-schemas is in principle the same.



```

<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns="http://www.interlis.ch/INTERLIS2.2"
  targetNamespace="http://www.interlis.ch/INTERLIS2.2"
  elementFormDefault="qualified" attributeFormDefault="unqualified">
  ....

  <xsd:complexType name="railML.infrastructure.singleTrackData">
    <xsd:sequence>
      ....
      <xsd:element name="trackTopology" minOccurs="0">
        <xsd:complexType>
          <xsd:sequence>
            <xsd:element name="railML.infrastructure.trackTopologyType"
              type="railML.infrastructure.trackTopologyType"/>
          </xsd:sequence>
        </xsd:complexType>
      </xsd:element>
      ....
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="railML.infrastructure.trackType">
    <xsd:sequence>
      <xsd:element name="trackID">
        <xsd:simpleType>
          <xsd:restriction base="xsd:string">
            <xsd:maxLength value="20"/>
          </xsd:restriction>
        </xsd:simpleType>
      </xsd:element>
      <xsd:element name="type" minOccurs="0">
        <xsd:simpleType>
          <xsd:restriction base="xsd:string">
            <xsd:enumeration value="mainTrack"/>
            <xsd:enumeration value="secondaryTrack"/>
            <xsd:enumeration value="connectingTrack"/>
            <xsd:enumeration value="sidingTrack"/>
            <xsd:enumeration value="stationTrack"/>
          </xsd:restriction>
        </xsd:simpleType>
      </xsd:element>
      <xsd:element name="trackName" minOccurs="0">
        <xsd:simpleType>
          <xsd:restriction base="xsd:string">
            <xsd:maxLength value="20"/>
          </xsd:restriction>
        </xsd:simpleType>
      </xsd:element>
      <xsd:element name="mainDir" type="railML.dirValidityType"
        minOccurs="0"/>
      <xsd:element name="singleTrack" minOccurs="0">
        <xsd:complexType>
          <xsd:sequence>
            <xsd:element name="railML.infrastructure.singleTrackData"
              type="railML.infrastructure.singleTrackData"/>
          </xsd:sequence>
        </xsd:complexType>
      </xsd:element>
    </xsd:sequence>
  </xsd:complexType>
</xsd:schema>

```

Figure 5: XML-schema for the INTERLIS-XML transfer format corresponding to the XML-schema for railML© of figure 1.



- INTERLIS-XML uses mainly XML-elements to encode conceptual attributes whereas railML© works primarily with XML-attributes. XML-elements are easier extendible whereas XML-attributes could be faster treatable.
- In INTERLIS-XML all inherited elements are explicitly included in the specialised class whereas in railML© they are only cited by the keyword `extension base`. Because this citation possibility is not supported by all XML-tools, the designers of INTERLIS-XML decided to use the inclusion.
- INTERLIS-XML uses at different places `mininclusive/maxinclusive` value instead of `total/fraction` digits. This provides a more exact basis for range checks.
- From the point of view of range checking maximal text lengths – as necessary for the actual INTERLIS version – are an advantage too. But in later INTERLIS versions this restriction will no more be mandatory.

There exist several possibilities to overcome these differences between INTERLIS-XML and railML©.

- A 1:1 processor railML© ↔ INTERLIS-XML can be developed.
- The INTERLIS compiler can be completed: Beside the format descriptions for INTERLIS-XML, for ITF and for GML the XML-schema for railML© could as well be produced according to a given conceptual schema.
- Finally the XML-schema of railML© can be described as exactly as possible on the conceptual level by UML/INTERLIS. The corresponding XML-schema for INTERLIS-XML is automatically derived and used as new XML-schema for a revised railML©.

This shows that there exist different possibilities to open the door to the advantages of the model driven approach for the railML© initiative.

4 Exchange of railway data across country borders

One result of our demonstrator work in the GeoRail project was the restructuring of German (DB) track data to a (reduced) Swiss track data structure including the coordinate transformation to ETRS. Then we arrived to commonly visualize the track geometry based on DB- and SBB-data at the border between Germany and Switzerland in the region of Schaffhausen ([1], chapter 5.9 to 5.11).

Now we would like to transfer data between the relatively complex track element based model (called TrEI) and a pure network model (called PureNet). Phase (A) of the MDA provides the two graphical conceptual schemas in UML as figure 6 and figure 7. The complexity of the TrEI model has to do with the fact, that track lines are subdivided into ordinary track elements and into switch track elements. The ordinary track elements are linked by the start- and end-point associations to switch points. But the switch track elements are directly associated to the switches. A switch object itself is a composition of 3 or 4 switch point objects (as can be seen in figure 6). Therefore it will not be easy to collect all track elements of a track line in the TrEI model.



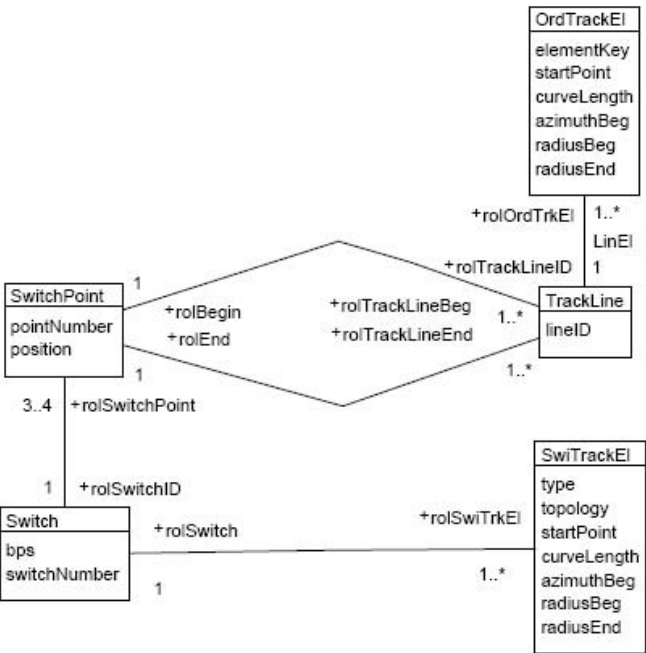


Figure 6: Conceptual schema TrEl as UML-Diagram.

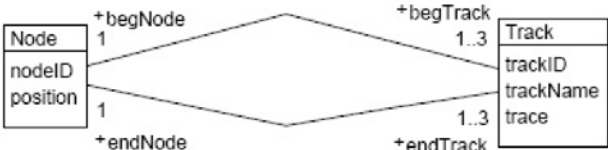


Figure 7: Conceptual schema PureNet as UML-Diagram.

The PureNet model on the other side is much simpler, because it consists only of the classes node and track. The nodes correspond to the (TrEl) switch points. The tracks are the glued collections of (TrEl) track elements (ordinary and switch ones) between two nodes described by the attribute `trace`, which is of type `POLYLINE`. Additional attributes have to be added in the PureNet classes to not loosing the switch track element properties of the TrEl model.

The corresponding textual conceptual schemas are using the textual conceptual schema language INTERLIS 2. The precise specifications of the data types are now given for every attribute. In the two models we have two different types for modeling the same geometric property. For track elements (of ordinary and of switch type) the geometry is defined by the coordinates of the start point, the element length along the curve, the azimuth of the curve tangent in the starting point, the radius in the starting point and the radius in the end point.



With PureNet the whole trace of a track is defined by the data type `POLYLINE` which has straight line segments, circle arcs, and clothoid segments (more types exist) as possible connection geometries of a single part (track element) of the whole track.

```

MAPPING MODEL TrEl2PureNet =
  TRANSFORMATION GROUP Group1 =
    SOURCE { TrackLine, TrackEl, ASSOCIATION LinEl}
    TARGET { Track }
    MAPPING {
      RULE TrkLine2Trk { Track := TrackLine; }
      RULE Glue { setPOLYLINE(LinEl, Track.trace); } }
    END Group1;
  END TrEl2PureNet;

```

Figure 8: Mapping two models.

In phase (D) the mapping from the TrEl model to the PureNet model has to be defined. Doing this in figure 11 we use an INTERLIS-like definition language for map description (see also Gnägi et al. [3]). At the moment this mapping language needs to be translated into one of the implemented mapping systems mentioned in section 2 (like ICS or FME). The example shows, how by a program knowing the two treated models an ordered composition of curve elements can be translated into a single attribute `trace` of type `POLYLINE`. The declarative definition of the maps allows different techniques or implementations.

As described in section 2 the semantic transformation just defined needs not only the conceptual schemata of the models in question and the mapping definition but in addition also the data to be transformed in a standard format corresponding to the model. Because the data were in a special proprietary format an own 1:1 processor had to be programmed in phase (C). Its result has been used as input to the semantic transformation, whose output was again in the standard format. With a checker tool the result of the 1:1 processor as well as the result of the semantic transformation can be tested against the conceptual schema. With this quality checks ends the restructuring process starting with the system TrEl and resulting in data according to the system PureNet. The whole chain of system-neutral tools used without modification of the start and the target GIS will be shown in a live presentation at AGILE 2006.

5 Further work

The differences in the modelling principles of railML© (data type nesting) and of the MDA (class and association graph) need a deeper analysis. The linking possibilities between the two methods mentioned at the end of section 3 should be seriously compared to evaluate, which strategy provides the optimal synergy effect with the minimal effort. By looking at the details of railML© we asked us the following question: Can the railML© data structure be simplified without loss of information? We would like to answer this question. To gain more experiences about methods and structures other railway application areas beside



infrastructure will be treated with the MDA. Clearly there is an urgent need to implement the conceptual mapping language used in section 4.

6 Conclusions

We have presented two approaches to railway data and their exchange. On one side is the detailed collection of huge railway knowledge at the XML-format level by the railML© initiative. On the other side offers the system- and format-independent model driven approach the obvious advantages e.g. of automatic quality testing and of easy data restructuring without a need for system restructuring. A successful combination of the two methods is conceivable and seems us to provide a solid basis for a sustainable interoperability and integrability of railway data across Europe.

References

- [1] Engel, T., Barbu, G., Gnägi, H. R., Lahr, B., Müller, S., Robert, D., Stahel, N., Winter, P., *Gleisbau und Gleisunterhalt auf der Basisabsoluter Koordinaten. Schlussbericht des UICProjektes Georail*. SBB, Bern, 15.5.2005
- [2] Engel T.: Barbu, G., Gnägi, H. R., Winter, P., Georail, presentation, NAVSAT, Geneva Switzerland, 23rd June 2003.
- [3] Gnägi H.R., Morf A., Staub P., Semantic Interoperability through the Definition of Conceptual Model Transformations. *Proc. of the 9th AGILE International Conference on Geographic Information Systems*, to be published 2006.
- [4] Hofmann R.: *Modellbasierte Datendrehscheibe für die SBB*, Bericht Vertiefungsblock, ETH Zürich D-Baug IGP-GF, 2003
- [5] INTERLIS 2, reference manual, user manual, tools. www.interlis.ch
- [6] railML© initiative, www.railml.org
- [7] Stahel N., Gnägi H.R., Georail Demonstrator. Presentation, UIC-Georail-Phase II: CNTD and ETRS, Workshop Paris, 15th September 2004



This page intentionally left blank

Formalisation and simulation of operating rules using coloured Petri nets

O. Lahlou¹, P. Bon¹ & L. Allain²

¹*INRETS, Villeneuve d'Ascq, France*

²*ISEN, Lille Cedex, France*

Abstract

The operational rules safety assessment of rail systems is a key element of the rail success to improve the competitiveness. In this context, the goal of this paper is to present a process for a safety analysis of operating rules. The first step of this process describes the operating rules and their formalisation using coloured Petri nets by means of a systematic method. The second step concerns the simulation of the resulting nets to check their properties.

A group of ERTMS (European Rail Traffic Management System) operating rules, stemmed from HEROE (Harmonization of European rail Rules for Operating) project concerning the departure of a train will be used as an example. These rules are “Departure”, “Train preparation”, “Permission of train movement authority” and “Written orders”.

Starting with a reminder of our systematic method of description based on coloured Petri nets (CPN), we show that the main advantage of this method is to standardize the rules description.

Then, the properties of the resulting Petri nets are checked with appropriate tools, with a view to verifying liveness and to detecting possible deadlocks.

Finally, we think that the process of formalisation and simulation will allow to check the consistency and the integrity of operating rules.

1 Introduction

From the perspective to harmonize the European railway system, ERTMS (European Rail Traffic Management System) [1], which is the new European standard for train control systems, has been developed and must be implemented for upgrades of high-speed and conventional lines in accordance to the European Commission laws.



In this way of harmonization, operating rules have an important role. Indeed, common operating rules in European railway will allow interoperability without constraint like knowledge of full regulations of each country for the driver or the presence of all country control systems on board the train.

Operating rules are regulations that drivers, railway staff and signalmen must apply to guarantee safety railway traffic. They define what must be done (or not) and the conditions to be applied.

Our aim in this paper is to present an approach for the verification of operating rules by the use of formal methods. Our case study depends on a group of ERTMS operating rules, stemmed from HEROE (Harmonization of European rail Rules for Operating project) [2], concerning the departure of a train. The first step of this approach describes the operating rules and their formalisation using coloured Petri nets (CPN) by means of a systematic method. The second step concerns the analysis of the resulting nets by means of simulation (which is equivalent to program execution and program debugging) or by means of more formal analysis methods to check the operating rules properties.

2 Context and problematic

Each country owns its operating rules. This implies that the traffic of trains between those countries is very difficult. To overcome this problem, common operating rules for the whole European railway network are the solution to allow a safety interoperability.

In this way, the HEROE project was launched in 1998 with the main objectives of harmonizing and assessing rules and regulations for the new ERTMS control-command system in normal and degraded modes. These common rules are mandatory upon all national railways undertaking using ERTMS. They will have to incorporate these rules, as applicable, within their national rules of operation.

The assessment approach developed during the HEROE project is based on the qualitative and quantitative analyses of ERTMS operating rules by using a common operational diagram to represent each rule.

The former analysis named qualitative assessment, checks that the safety will be fully effective if the rules are applied. The latter performs the quantitative analysis focused on the rules identified by the former as very critical. In fact, the quantitative assessment consists in evaluating the probability that the operators made errors in applying the rules (due to their complexity or for human errors) and evaluating the relevant consequences from the safety viewpoint.

This work identified some problems of inconsistency and incoherence, so it recommended to refine some rules [3]. Unfortunately this approach requires experts of railway domain to analyse the rules. Even if rules formalisation requires experts, we propose the use of formal methods in order to prove operating rules.

The HEROE works allowed us to identify a particular semantics of rules. Indeed, we can find events, conjunctions, disjunctions, sequential and simultaneous actions. This semantics reminds the ECA rules notably used for the modelling of active databases [4].



ECA rules can be modelled with Petri nets [5, 6]. Thus, the Petri nets paradigm seems particularly relevant to describe the operating rules. Petri nets are a graphical and mathematical tool. They are a well established formalism for modelling and verifying concurrent and reactive systems. We invite the readers to refer to [7] for more explanation.

Nevertheless, simple place/transition Petri nets are insufficient because we have to model data transmission between events, actions and conditions. In order to answer this transmission problem, we use coloured Petri nets.

Coloured Petri nets (CPNs) [8] were introduced as an extension to Petri nets, allowing tokens to carry a data type and a value. The advantage of coloured Petri nets is the combination of the strength of Petri nets and the strength of programming languages. Indeed, Petri nets allow the description of the synchronization of concurrent processes, while programming languages provide the primitives for the definition of data types and the manipulation of data values.

We now present an example of operating rules description using a systematic method. It concerns a group of rules taken from the HEROE project.

3 Practical example

3.1 Group of rules

Our practical example concerns the group of rules applied for a safe departure of a train stemmed from the HEROE project (ERTMS/ETCS operating rules). It includes the rules “Departure”, “Train preparation”, “Permission for train movement” and “Written order”. Figure 1 shows the links between these rules.

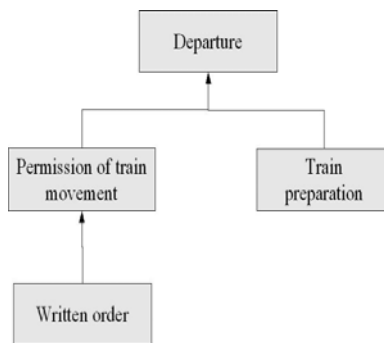


Figure 1: Relations between rules.

Figure 2 gives the text of the “Departure” rule. This rule requires five conditions to be performed. So, this is clearly a conjunction. We can also deduce that the rules for “Train preparation” (figure 3) and for “Permission for train movement” (figure 4) are called by the “Departure” rule. These relations between rules create



a link between their descriptions. Consequently, we use hierarchical coloured Petri nets to model the operating rules.

- The driver of a train is allowed to depart when :
 - he has information that conditions of “Train preparation” are met *and*;
 - he has information that the closing doors procedure has been completed and no passengers are in danger *and*;
 - the departure time is reached *and*;
 - he has received a “Permission for train movement” *and*;
 - the train is capable of clearing the platform.

Figure 2: Departure.

Figure 3 shows the “Train preparation” rule. A train must be prepared at the station (when it is composed) and in all cases when its composition or brake performance have been changed. This rule is also a conjunction.

- A train is prepared, when
 - all vehicles it consists of are coupled,
 - the train is equipped with train signals,
 - technical checks are successfully performed,
 - brake test is successfully performed,
 - brake performance is calculated,
 - train data are available,
 - data entry procedure has been performed,
 - information to check the integrity of the train is available.

Figure 3: Train preparation.

“Permission for train movement” rule is given in figure 4. We can see that signalman and driver actions are exclusive and only one of these should be executed at a time. All signalman and driver actions are obviously combine by disjunction operator.



- Signalman : when all required conditions are fulfilled, the signalman may permit a train movement by :
 - issuing a (full supervision or on sight) movement authority or
 - clearing a main signal or
 - issuing a written order
- Driver : If the driver receives :
 - a full supervision or on sight movement authority or,
 - a signal displaying a proceed aspect or
 - a written order
 he has permission to move his train.

Figure 4: Permission for train movement.

“Written order” rule (figure 5) is a safety related message issued by the signalman to the driver. It is used to convey an instruction that cannot be achieved by ERTMS. This rule is composed with parts that containing different semantics.

- Conditions :
 - train is stationary,
 - driver has identified his function, train number and location,
 - train has reached the last EOA in front of the section concerned.
- Content :
 - who issued it
 - from where it is issued
 - at what time and date
 - to which train it is intended
 - the instruction itself in an unambiguous manner
- Ways of transmission :
 - written orders may be transmitted physically written on paper, as text message or as a verbal message the driver must write it down.
- Permission of train movements :
 - the signalman must not permit a train movement until the written order has been received and confirmed by the driver.
- Precedence over DMI indication :
 - written orders take precedence over all relevant indications provided by the ERTMS DMI.

Figure 5: Written order.

Now we have four rules at our disposal in order to depict the systematic method of description presented in the next section.



3.2 Description method

In this section, we present the equivalent description of a rule referring to its semantics. Indeed, we give each rule semantics a description based on coloured Petri net. This systematic method of description allows to have a standardisation of rule translation, even if the description is done by different people.

We spotted that rules are based on events, conjunctions, disjunctions, sequential and simultaneous actions. Now we describe the translation method of these elements step by step.

- An event is represented by firing a corresponding transition of the CPN.
- Conditions combined by the conjunction operator may simply use a transition to synchronize all fulfilled sub-conditions as shown in Figure 6. This can be applied for instance to the “Departure” and the “Train preparation” rules.

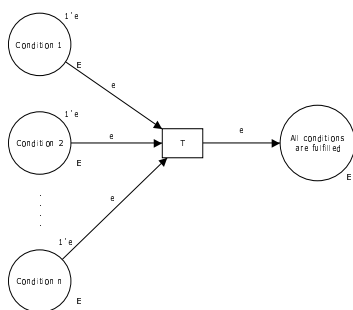


Figure 6: Conjunction.

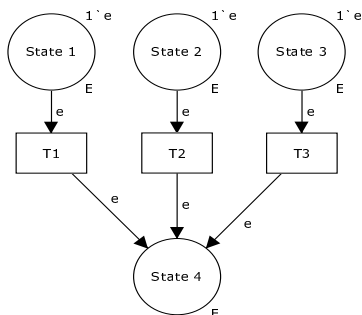


Figure 7: Disjunction.

- Conditions combined by the disjunction operator use multiple transitions as shown in figure 7. Each of them is related to one sub-condition. This translation assumes that all sub-conditions are exclusive.
- Sequenced actions are automatically described into a sequence of places and transitions as shown in figure 9.
- A rule called by another one, as for the “Train preparation” and “Permission for train movement” called by “Departure” rule, is described by a hierarchical net. An additional transition is added to depict the link (figure 8).

To understand the different annotations that we use in our models, we give here some details of CPN elements. We invite the readers to refer to [9] for more explanation.

- Each place has an associated type (colour set) determining the kind of data that the place may contain. In our example, we use E as colour. It is also possible to use complex type with several fields containing string, integer and other complex types.



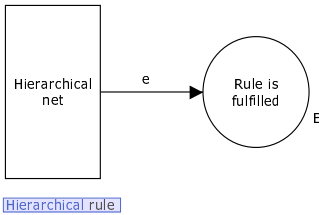


Figure 8: Hierarchical net.

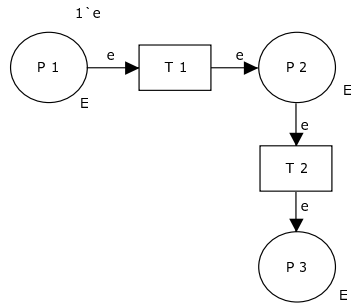


Figure 9: Sequenced actions.

- The marking of a place consists of the number of tokens this place contains. It is indicated as $1'e$ in our examples. Each token carries a value (colour), which belongs to the type of the place on which the token resides.
- To fire a transition, we need to assign data values to the variables occurring in the arc expressions. Then, we evaluate the arc expression to remove and add tokens respectively from input and output places. Either variable occurring in the arc and input-output places must have the same colour.

To complete the presentation of this systematic method of description, figure 10 shows the result of its application for the “Departure” rule (figure 2). To reach this aim, we used the conjunction translation to describe the five conditions listed on the rule. We also used two hierarchical nets to describe the call of “Train preparation” and “Permission for train movement” rules.

We presented here a systematic method of description of operating rules by means of coloured Petri nets. This method will allow a standardisation of coloured Petri nets describing operating rules. The use of hierarchical nets to describe a call of rule preserve the structure of operating rules and the link between them. It allows also to reuse Petri nets when the corresponding rule is called many times.

4 Discussion

CPN allows investigation of the system behaviour by making simulation of its model. Indeed, the simulation provides a great amount of detail, particularly for very large CPN models. Thus, in our case, we can follow the progress of operating rules execution with a view to observe message transmissions.

On the other hand, the state space method of coloured Petri nets makes it possible to validate and to verify the functional correctness of systems [10]. Indeed, with the state space method, it is possible to answer a set of analyses and verification questions concerning the behaviour of the system. Then we can verify the properties like the presence or not of deadlocks in the system. This can help us to detect scenarios which can affect the safety of railway system, using CPN-Tools for instance [11].



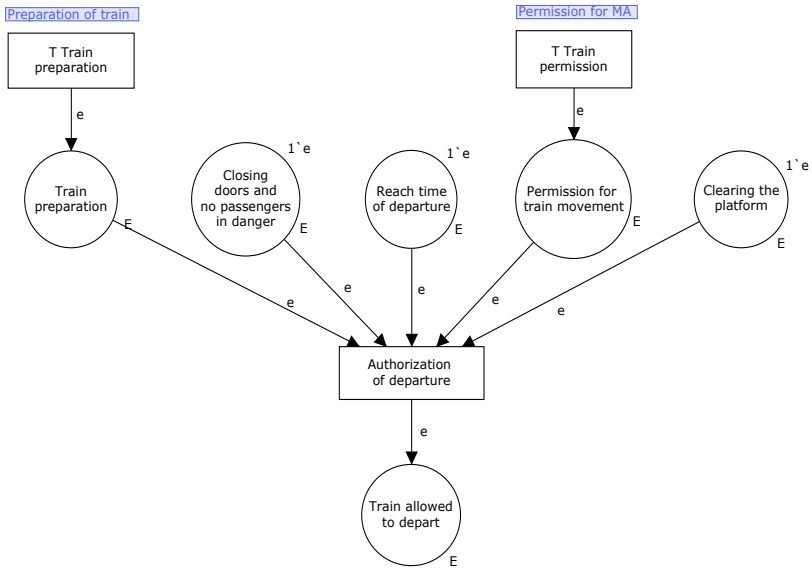


Figure 10: Complete CPN model for departure rule.

We can also use the work of [12] to translate the coloured Petri nets on an abstract B machine. Here the aim is to use the power of this method like the refinement and well-established proof tools [13].

5 Conclusion

We presented in this paper an approach to verify the safety of operating rules. We chose coloured Petri nets because it seems particularly relevant depending on two points. The former is the similarity of the semantics of operating rules and ECA rules which can be modelled with Petri nets. The latter is to have a set of analysis methods and simulation tools at our disposal. For instance, coloured Petri nets paradigm can help us to verify a deadlock situation during execution of operating rules or to detect some dangerous scenarios.

References

- [1] ERTMS, Functional requirements specification. **V4.29**, 1999.
- [2] HEROE project WP A, Operational principles and rules. *The ERTMS Users Group*, p. 95, 2001.
- [3] El-Koursi, E.M. & Kampmann, B., Qualitative and quantitative safety assessment of ERTMS operating rules. *CompRail VIII, Lemnos, Greece*, pp. 671–680, 2002.



- [4] Hanson, E.N. & Widom, J., An overview of production rules in database systems. *The Knowledge Engineering Review*, **8(2)**, pp. 121–143, 1993.
- [5] Allain, L. & Yim, P., Modeling information system behavior with dynamic relations nets. *JUCS*, **6(11)**, pp. 1109–1130, 2000.
- [6] Gatziau, S. & Dittrich, K.R., Detecting composite events in active database systems using petri nets. *RIDE-ADS*, pp. 2–9, 1994.
- [7] Murata, T., Petri nets: Properties, analysis and applications. *Proc IEEE*, **77(4)**, pp. 541–580, 1989.
- [8] Jensen, K., *Coloured Petri Nets: Basic Concepts, Analysis Methods, and Practical Use*, volume 1 of *EATCS Monographs in Computer Science*. Springer-Verlag, 1992.
- [9] Kristensen, L.M., Christensen, S. & Jensen, K., The practitioner's guide to coloured petri nets. *STTT*, **2(2)**, pp. 98–132, 1998.
- [10] Huber, P., Jensen, A.M., Li, Jepsen, L.O. & Jensen, K., Reachability trees for high-level petri nets. *Theoretical Computer Science*, **45(3)**, pp. 261–292, 1986.
- [11] CPN Group, University of Aarhus, Denmark, CPN tools home page. [Http://wiki.daimi.au.dk/cpntools/](http://wiki.daimi.au.dk/cpntools/).
- [12] Bon, P., *Du cahier des charges aux spécifications formelles : une méthode basée sur les réseaux de Petri de haut niveau*. Ph.D. thesis, (in French), École Centrale de Lille, 2000.
- [13] Abrial, J.R., *The B-Book: Assigning Programs to Meanings*. Cambridge University Press, 1996.



This page intentionally left blank

From UML to B – a level crossing case study

J.-L. Boulanger¹, P. Bon² & G. Mariano²

¹*Universite de Technologie de Compiègne, HEUDIASYC,
Compiègne cedex, France*

²*INRETS, ESTAS, Villeneuve d'ascq, France*

Abstract

The goal of this paper is to show how it is possible to combine the advantages of Unified Modelling Language (UML) and of the *B* method in order to design safety applications. We investigate how the Unified Modeling Language (UML), can be used to formally specify and verify critical railways systems. A benefit of using UML is its status as an international standard (OMG) and its widespread use in the software industry. *B* is a formal method for the incremental development of specifications and their refinements down to an implementation. In the railway critical software domain, safety requirements are obviously severe. It is very important to keep requirements traceability during software development process even if the different used models are informal, semi formal or formal.

Keywords: B method, formal development, level crossing, software verification.

1 Introduction

In spite of progress carried out in software development, designing a complex system while respecting its safety requirements, remains very hard. During the critical software development process, safety and security requirements must be traced from informal specification to code generation. So we need to trace them in the different models: informal, semi formal or formal ones. We present a new method here to transform a semi formal modelling to a formal specification which enables them to be traced. This method will be applied to a railways case study, where safety requirements are very strict. We study a level crossing case study taking into account French particularities. This article is made up 3 parts. Firstly, we describe the case study. In the following part we present the principles of UML and we give a part of semi formal modelling of the level crossing. In the last part



we give a short presentation of the B method, and then we describe a part of the formal model generation from the semi formal one.

2 The case study : level-crossing

To illustrate our approach, we will choose to design a level crossing. This example is inspired by [8]. The term level crossing, in general a crossing at the same level, i.e. without bridge or tunnel, is especially used in the case where a road crosses a railway; it also applies when a light rail line with separate right-of-way crosses a road; the term “metro” usually means by definition that there are no level crossings. Firstly, a single-track line, which crosses a road in the same level, is modelled. The crossing zone is named danger zone. The most important security rule is to avoid collision by prohibiting road and railway traffic simultaneously on level crossing]. The railway crossing is equipped with barriers and road traffic lights to forbid the car passage. Two sensors appear on the railroad to detect the beginning (train entrance) and the end (train exit) of the level crossing protection procedure. The level crossing is not in an urban zone this implies a sound signalisation. Traffic lights consist of a single flashing red light. When they are switched off, road users (drivers, pedestrians,) can cross. In the other case, the level crossing is closed and railway traffic has priority. At any time, guards can control the level crossing. In this case, the guards must do all they can to ensure the level crossing safety.

3 Semi-formal specification

3.1 UML

Born from the different object methods, like OMT or Booch & Jacobson, and normalised by the Object Management Group, UML has now become a standard to model systems. The UML notation makes it possible to model an application according to an object view. 9 different diagram types make this modelisation. Each diagram allows a particular view of the system. The reader interested by more details in syntactic and semantic aspects can refer to the reference guide of UML [1]. Even if UML notation is a language in which models can be represented, it does not define the making process of these models. Nevertheless, several dedicated tools have strengthened the popularity of UML. These tools allow graphic notation and partial generation of the associated code and documentations. The UML notation is known by most computer scientists and is now used in several domains. Using UML class diagrams to define information structures has now become standard practice in industry. Recently, the critical application domains have used the notation and several questions exist around this use. By providing a rigorous semantic for a subset of UML and a systematic translation to the B method, we aim to propose a method, called **UML2B**, for developing safety critical system.



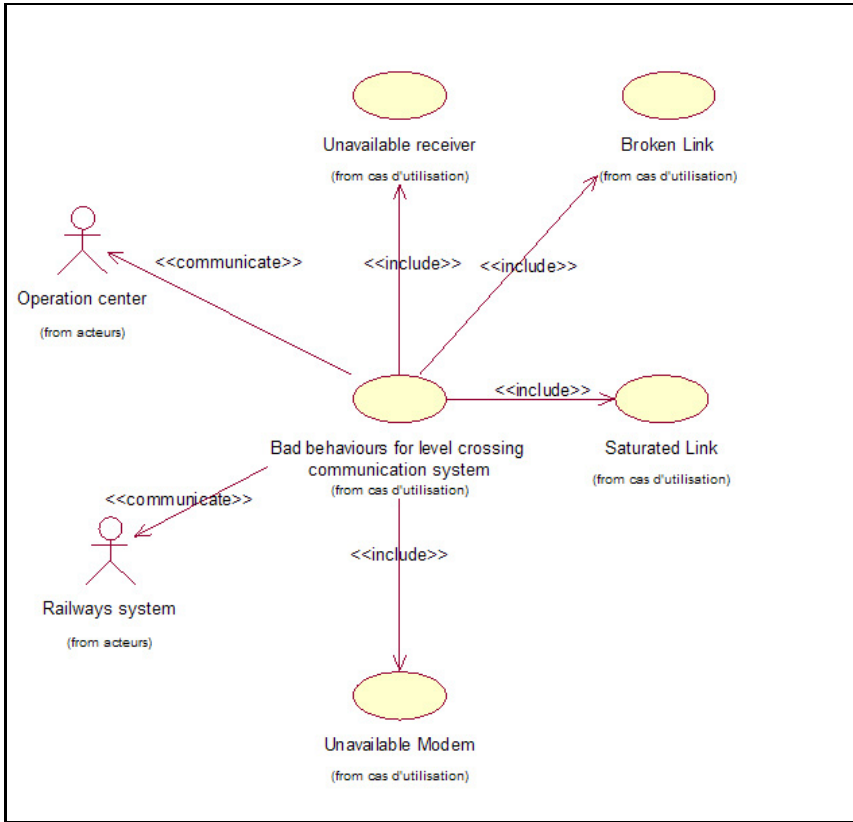


Figure 1: An usecase diagram.

3.2 Application

3.2.1 Environment and its interactions on the system

We describes the interaction between the physical environment and the level crossing system by some usecase diagrams (see an example in figure 1). An usecase diagram describes and traces the functional requirements of the system and describe how the system can and will be used. The usecase diagram gives an overview of the model. In UML, objects communicate with and send messages to each other. We uses sequence diagram (see figure 2) to describe how objects interact and communicate with each other. The focus is time.

3.2.2 Architecture of level crossing system

The level crossing system architecture is describe in the class diagram (see figure 3), which describes the relationships between classes and shows the logical view of a system (static view).



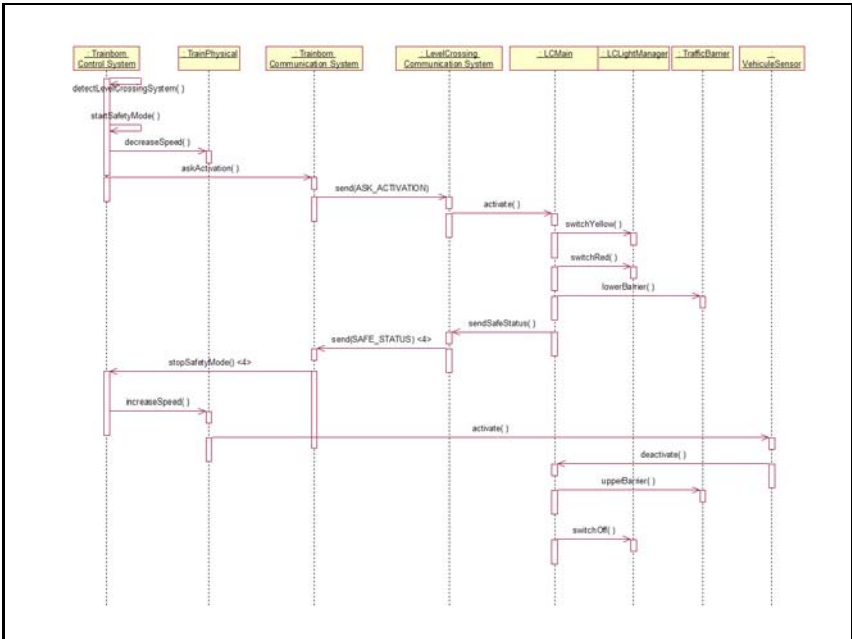


Figure 2: A sequence diagram.

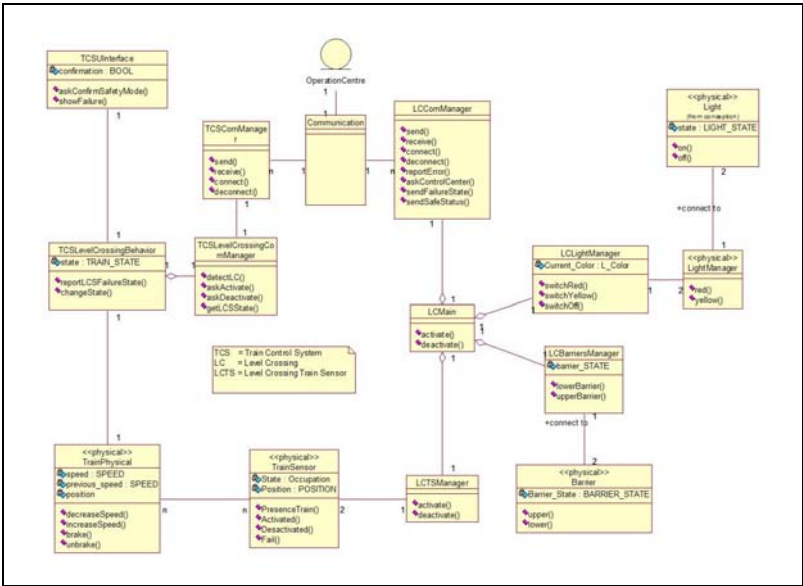


Figure 3: Class diagram.



3.2.3 Behaviour

In the next part of the paper, among the different possible diagrams, we'll use the state diagrams particularly adapted to reactive system modelling. A state is a condition in an object life while it satisfies some conditions, runs some actions or waits for some events. There are two special states: initial state and end state. The Initial State is the state of an object before any transition. End States marks the destruction of the object whose state we are modelling. An event is a particular occurrence that can trigger a transition from a state to another one. A state diagram can represent the system behaviour. This type of diagram represents finite state automaton, under a graphical representation, linked by oriented arcs describing transitions. Statechart diagrams [6, 7], also known as State diagrams, are used to document the various modes ("state") that a class can go through, and the events that cause a state transition. The state-transitions graph formalism is not a UML innovation. It has often been employed in other contexts and a large consensus, from David Harel's works, exists around this notation. It introduces the description of possible sequences of states or actions which can occur to an element during its life. Such sequences arise from element reaction to discrete events.

3.2.3.1 Level crossing control system. Gates, lights and bell are managed by the level crossing control system (Figure 4). It will be activated as soon as a train approach the level crossing. The activation is produced by the railway traffic. The point, from which activation is produced, is called start of danger zone. A time allowed is needed between the activation and the arrival of the fastest train into the level crossing. This time, named train on line period, depends on maximal train speed into the level crossing. The moment the control system is activated, a sequence of orderly action is launched by the control system in order to empty the level crossing in time and protect it from road traffic. First, the lights switch on in order to stop the road traffic. At the same time, the bells ring (in a non urban zone). After a notice time, the barriers pull down. If there is no problem during the pulling down (i.e. pulling down made in the maximal temporal limits), the system is in safe mode and the train will have the right to cross. After the train crossing, the barrier pull up and the lights and bells switch off. The control system can at any time go to manual mode. In this mode a guard manages the level crossing security. This mode can be activating when a material problem occurs (no pulling down, no light switch on,).

3.2.3.2 Embedded system. The embedded system runs several actions when the train comes near a crossing level (Figure 5). When the train pass the start of danger zone, the embedded system asks to the control system an acknowledgement (ack), the embedded system gets into stand by and begin to brake in order to pull down the barrier in time. After this notice time, the control system sends its state to embedded one. If the level crossing is in safe mode, the embedded system stops the braking and restarts with its initial speed. An end-crossing sensor detects the train exit and starts the barrier pull up and the lights switch off.



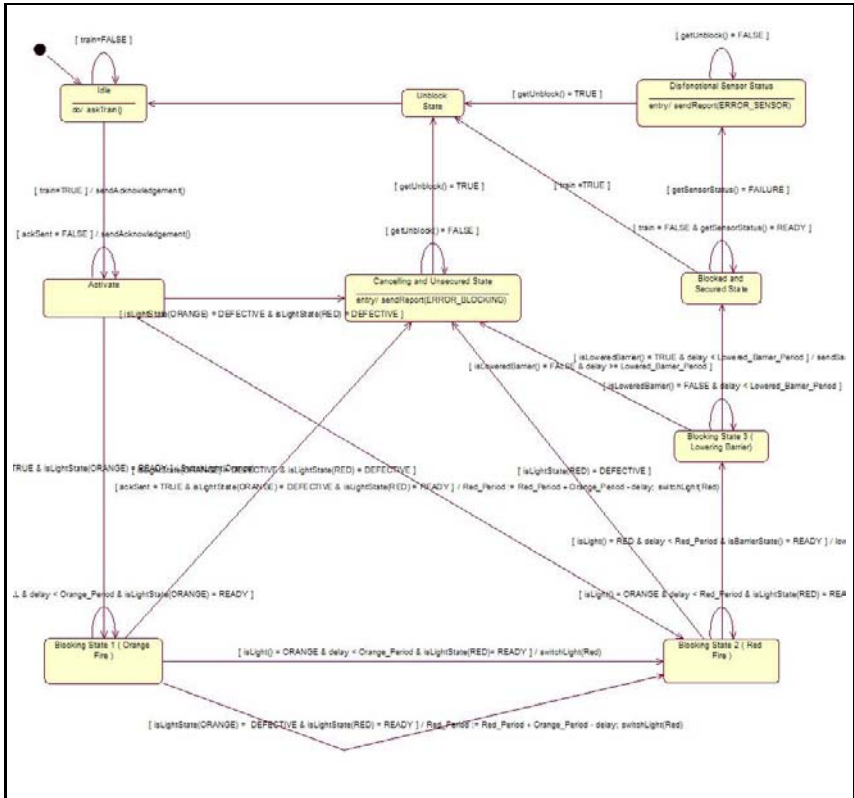


Figure 4: State transition diagram of level crossing control system.

4 Formal specification

4.1 The B method

The *B* method seems to be the most appreciated formal method in industrial world for railway critical software development such METEOR (see [3] or [5]). In fact, *B* method allows building a gate between mathematical modelling and informatics realisation. The *B* method due to J.R Abrial [2] is a formal method for the incremental development of specifications and their refinements down to an implementation. It is a model-based approach similar to **Z** and **VDM**. The *B* method covers all the software development process through a series of proved refinement steps. The software design in *B* starts from mathematical specifications(set description, first order logic and substitution). Little by little, through many refinement steps [9], the designer tries to obtain a complete and executable specification. This process must be monotonic, that is any refinement has to be proved coherent according to the previous steps of refinement. The *B*

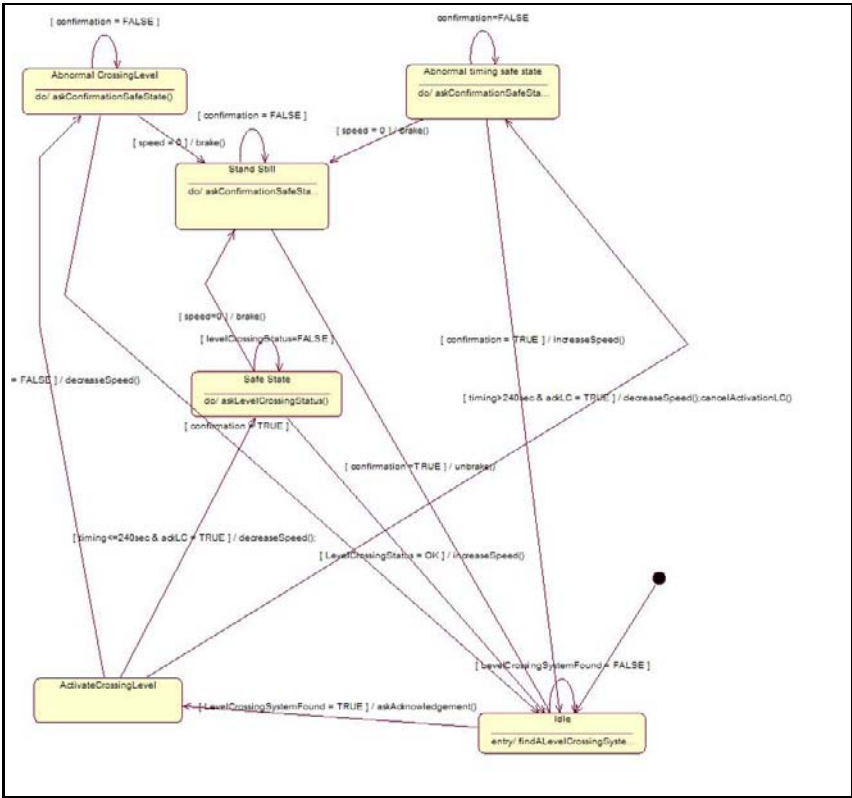


Figure 5: Embedded system running.

method allows generating a secured and proved code from formal mathematical expressions. The **abstract machine** is the basic element of a *B* development. It encapsulates some state data and offers some operations. The *B* tool allows us to generate automatically the proof obligations for each abstract machine. Generally speaking, the proof obligations will be more and more complex as concrete details are introduced. Then these proof obligation are discarded either automatically for the simple ones or in cooperation with the designer for the complex ones. So, at the last refinement called the implementation, we obtain a secure software which does not need to be tested. At this low level stage, it may be easily translated automatically to a programming language.

4.2 *B* specifications generated from UML diagrams

This section presents the *B* specification generated from UML diagram for the Level Crossing problem. For space constraints, we do not describe all the specifications. We fully presented the *B* specification generation processus in [4].

```

MACHINE TCS_LevelCrossingBehavior_0
SETS STATES = {Idle, SafeState, ActivateCrossingLevel, StandStill,
               AbnormalTimingSafeState, AbnormalCrossingLevel}
CONCRETE_VARIABLES current_state
INVARIANT current_state : STATES
INITIALISATION current_state := Idle
OPERATIONS
change_state =
BEGIN
current_state : (
  current_state : STATES
& ((current_state$0= Idle) => current_state: {ActivateCrossingLevel, Idle})
...
& ((current_state$0= StandStill)
  => current_state /: {SafeState, AbnormalTimingSafeState})
)
END
END

```

The previous *B* abstract component, called *TCS_LevelCrossingBehavior_0*, we introduced all states in set called *STATES* and an operation called “change_state”. This abstract component is not deterministic since we use the operator *list_var : (predicate)* in the OPERATION clauses. This operator indicates that the list of variable become such that the predicate is true. In the next abstract machine, we implemented the abstract component “*TCS_LevelCrossingBehavior_n*”. This abstract machine introduces the dynamic of the state diagram.

```

IMPLEMENTATION          TCS_LevelCrossingBehavior_n
REFINES                  TCS_LevelCrossingBehavior_0
INVARIANT (current_state = StandStill) => (brake = TRUE)
  & (current_state /= StandStill) => (brake = FALSE)
INITIALISATION          current_state := Idle
OPERATIONS
change_state =
CASE current_state OF
  EITHER Idle THEN
    VAR bb IN
      bb <-- detectLevelCrossingSystem;
      IF ( bb = TRUE )
      THEN
        BEGIN current_state := ActivateCrossingLevel
          ; AskAcknowledgement
        END
      ELSIF ( bb = FALSE)
      THEN current_state := Idle
      END
    END
  END
...
END
END

```

5 Conclusions

With these two main parts, Semi formal modelling with UML and formal specification with *B*, this works show the first step that allow developing a completely computerised level crossing. This application will be totally proved and will guarantee an optimal safety. Of course the real development must be done, but



with some add-on, the method presented here allow considering such development. Such an example of future work, safety invariants can be derived from hazard analysis and coded in UML by using OCL constraints attached to classes or sets of associations to specify safety and operational invariants of reactive systems in a concise manner. We will translate these UML constraints (OCL) into invariants and control rules in the B model. The main difficulty to specify railway case study is the less of harmonisation between the different European systems. The level crossing modelling presented here give a first step to a computerised management of level crossing.

References

- [1] Unified modelling language version 1.4. Technical report, OMG, 2001.
- [2] Jean-Raymond Abrial. *The B Book - Assigning Programs to Meanings*. Cambridge University Press, August 1996.
- [3] Patrick Behm, Paul Benoit, Alain Faivre, and Jean-Marc Meynadier. METEOR : A successful application of B in a large project. In *Proceedings of FM'99: World Congress on Formal Methods*, pages 369–387.
- [4] Philippe Bon, Jean-Louis Boulanger, and Georges Mariano. Semi formal modelling and formal specification: UML & B in simple railway application. In CNAM-Paris, editor, *ICSSEA 2003*, December 2-4 2003.
- [5] C. DaSilva, B. Dehbonei, and F. Mejia. Formal specification in the development of industrial applications: The subway speed control mechanism, 1991.
- [6] David Harel. On visual formalisms. *Communications of the ACM*, 31(5):514–530, 1988.
- [7] David Harel and al. On the formal semantics of statecharts. In *Proceedings of the 2nd IEEE Symposium on Logic in Computer Science*, pages 54–64. IEEE Press, 1987.
- [8] L. Jansen and Eckehard Schneider. Traffic control systems case study: Problem description and a note on domain-based software specification. Technical report, Institute of Control and Automation Engineering, Technical UNIVERSITY of Braunschweig, 2000.
- [9] C. Morgan. *Deriving programs from specifications*. Prentice Hall International, 1990.



This page intentionally left blank

Section 6
Converting metros to
driverless operation
(special section organised
by A. F. Rumsey)

This page intentionally left blank

The feasibility case for converting existing heavy metro systems to driverless operation

P. Thomas

Parsons Group International Ltd, London, UK

Abstract

Bill Gates once said that ‘The first rule of any technology used in a business is that automation applied to an efficient operation will magnify the efficiency. The second is that automation applied to an inefficient operation will magnify the inefficiency.’

The concepts and principles of driverless metro systems are well established, especially when applied to Automated People Movers and new build Metro Systems. Driverless operation can deliver significant operational and service benefits over conventionally operated lines. With the establishment of Driverless systems and the underlying enabling technologies the business case for converting an existing metro system to driverless operation is becoming increasingly feasible. To this extent many existing metro operators are either considering, planning or implementing driverless operation.

When considering converting an existing metro operation from a conventionally staffed operation a number of factors need to be assessed by the operator to evaluate the feasibility of the endeavour. These include but are not limited to: the operational and service benefits required from the system; the safety risk and subsequent control measures; the public perception and acceptance of the system; systems integration and migration from the existing to new system.

This paper will assess the principle factors that need to be addressed in the feasibility case for conversion of a heavy metro system to driverless operation. This paper is not setting out to provide an economic evaluation of such a scheme. However it *is* suggesting that it is possible to construct a feasibility case to convert an existing heavy metro system to driverless operation by considering the issues discussed in this paper.

Keywords: driverless operation, systems integration, systems engineering, automation, train control systems.



1 Limitations with conventionally operated automatic train operation

Conventional operations constrain the service offered by a metro operator in a number of key areas. With conventional operation the service is normally tied to the crew roster which makes it difficult to change the service pattern to respond to events on the network such as special events, perturbation recovery and schedule reformation. With the staff tied to the cab they are effectively unable to perform any other functions other than operating the train. Successful realisation of the benefits that are enabled through driverless operation requires a change to the operation of the railway. The key constraints and operational changes through driverless operation are summarised in Table 1.

Table 1: Summary of constraints.

Conventional Operation	Driverless Operation
Service requires a train driver located within a cab	Service capable of operating without staff, use of staff for operational benefit.
Drivers manage the Platform Train Interface (PTI) and despatch trains	Automatic management of PTI, door control, platform dwell time and train despatch. Optional manual override of door control for failure / dwell management
Drivers are able to carry out remedial actions	Remote diagnostic / reset facilities from the Control Centre
Perturbation recovery constrained due to crew plans linked to train schedule, staff availability and existing recovery point.	Perturbation recovery improved through operational flexibility via automatic-turnaround facilities and train service completely decoupled from staff plan.
The need for the driver to be on the trains as they come into and out of service constrains the flexibility of operating staff.	Operational staff can join train anywhere along the line and allows for increased staff flexibility.

2 Key safety concepts and safety risks

The introduction of driverless operation clearly impacts the overall safety of the railway. In some areas this can be seen as an increase in safety through reduction in human error and in other areas it is a reduction in the control measures afforded by the driver in the cab. This change in the safety of the railway, and the necessary control measures needs to be assessed as part of the business case for the project. Of particular importance is the extent to which implementation of these controls to an existing railway impacts the effectiveness of these safety measures.



The effectiveness of control measures and proven systems applied to comparable systems to control the identified driverless operations hazards need to be evaluated in accordance with a recognised risk assessment methodology such as that defined in EN 50126. Where appropriate, modified or new safety control measures need to be introduced to reduce the hazard to a tolerable level. This analysis results in an understanding of the 'delta' required to mitigate driverless operation hazards and materialise potential safety benefits over and above the current operational situation.

Using this approach a hazard log and risk ranking can be created for the driverless system. Table 2 illustrates the top three hazards arising from the evaluation of the safety risks of converting a conventionally operated metro system to driverless operation.

Table 2: Top hazards.

Top Event	Hazard	Possible Mitigation Options
Platform / Train Interface	Person On track	Platform Edge Doors
		Fall detection and prevention
		Minimise mid car gaps
		Passenger awareness
	Passenger Dragged by train	Platform Edge Doors
		Trapped passenger detection
		Operational staff at high risk areas / times
		Passenger awareness
Collision	Object on track	Object prevention
		Object detection
		Object deflector on train
	Infrastructure failure	Infrastructure failure detection
		Speed restriction
Ventilation in tunnel sections	Passengers trapped	Maintain air quality
		Train movement ability maintained or restored
		Passenger communications
		Remote driving
		Bi-directional signalling
		Evacuation to place of safety

2.1 Platform Train Interface (PTI)

Control measures used to ensure the safety risks associated with the PTI operation is tolerable include (as appropriate): The use of Platform Edge Doors (PED) fitted to the platforms to prevent access to the track area. PEDs will also provide an improved safety benefit over and above that of conventional



operation by eliminating falls into the track area. The use of a Platform Track Intrusion Detection System (PTIDS) which prevents a train from entering the platform upon detection of egress into the track area. The use of 'Smart' doors which provide improved detection and release of obstacles trapped between the closing edges of the train doors.

2.2 Collision with objects

Control measures used to ensure the safety risks associated with the collision hazard are tolerable include (as appropriate): Prevention of access to the track through the use of security fencing along the perimeter and over-bridges; the deployment of CCTV and / or other technologies at strategic locations (trespass hot spots). Use of 'Sweep' trains at the start of service, utilising additional vigilance to prove that the track is clear of obstruction. Obstacle deflection by the front of the train (such as 'cow catcher') to reduce the consequence of a collision with an object.

2.3 Ventilation in tunnel sections

Trapped passengers are generally defined as those in a train that has been stopped between platforms for more than 30 minutes. The passenger perception of a train stuck in a tunnel is worse than one stuck on the surface, furthermore with small bore tunnels with no side access, passenger panic and thermal affects rise with time, temperature and crowding. In driverless operation where a member of staff may not be on board the primary control measures are to maintain train movement, air quality and ensure good passenger communications. Passenger self detrainment through the end of the train is a last resort.

3 Customer perception

Customer research has found that Customer awareness of driverless systems tends to be minimal; some users of existing driverless and conventional Automatic Train Operation (ATO) systems are not aware that these trains are driven automatically. At first, most rejected the concept of Underground trains without drivers and said it was imperative for someone to be in charge of the train. Once customers are aware of existing systems many are persuaded that technology for driverless operation not only exists but is tried and tested. Most come to the conclusion that driverless operation will not be an issue, as long as safety worries are addressed. They also expect that such a development will be part of a larger modernisation programme, to include updated trains and stations, and in this context driverless operation becomes more acceptable.

Customers still, however, want a member of staff on trains, who ideally would be able to take over and drive in an emergency. Security is key, and the absence of the presence of staff is a real cause for concern.



4 Performance measurement and service benefits

When considering the performance benefits that can be achieved through driverless operation the primary measures are the improvement to the journey time and the change to the availability of the railway. Service benefits will also be realised through an increased visibility of staff capable of providing other service enhancing duties. The impact of driverless operation in these areas is discussed below.

4.1 Journey time

The improvements to journey time needs to consider the time to complete the journey under perturbation-free conditions and the additions to journey time caused by variations arising from: train running times; staff absences and errors; and minor equipment failures.

The key elements of the journey time measures that are impacted by driverless operation are discussed below. These are also summarised in figure 1 for a typical heavy metro system compared with the implementation of conventional ATO.

4.1.1 Dwell time management

Driverless operation will impact dwell management through a reduction in human intervention in the decision process to dispatch the train from the platform. Automatic door opening and closing will eliminate some of these inefficiencies but may also result in longer dwell times than necessary (as there is no driver to observe completion of the alighting and boarding process.) More regular dwell times enabled through automatic management of the platform dwell will lead to less variation in train frequency and provide a more regular service.

4.1.2 Automatic turn-around

The ability to automatically reverse the service without the need for human intervention or 'stepping back' will improve the turn-around time and allow an increase in train frequency. This can permit an improved service to be delivered with fewer trains than conventional operations and may also avoid costly infrastructure works to increase capacity at turn back locations.

4.1.3 Train capacity

Additional space is enabled by removal of cabs and possibly by using longer trains as the front and rear can extend beyond the ends of the platforms.

4.1.4 Decoupling train and staff schedules

Delays due to staff absences and unplanned personal-needs relief can be eliminated through driverless operation. There is more flexibility to optimise timetables when not constrained by staff duty periods. Train schedules can be adapted dynamically in response to perturbations or special events without being constrained by the availability of staff to operate the trains.



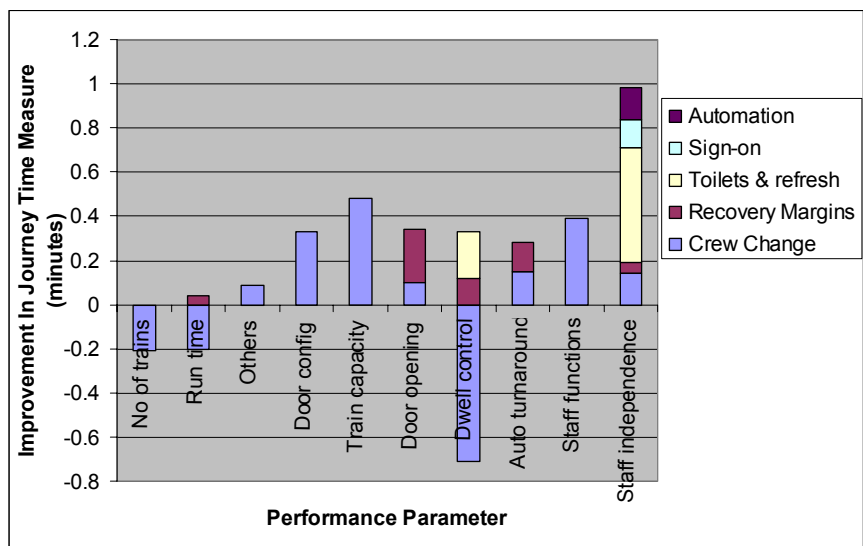


Figure 1: Improvement in Journey Time using conventional ATO as a baseline.

4.2 Availability

The availability of the railway will normally be expressed in terms of the loss of service through Service Affecting Incidents (SAI) and the time taken to recover from this Initial Delay to the service. The figure for SAI will normally be made up of asset failures and human error.

In this area driverless operation will also have a significant effect due to the introduction of new types of Service Affecting Incidents and modification of Initial Delays as there will not be a driver on-board to take remedial action. Remote fault diagnosis and rectification systems can be used to compensate for this effect. New degraded modes of operation will also be introduced by driverless operation which needs to be analysed. Some may adversely affect the availability in the absence of a driver but on the other hand, some train movements can be made more efficiently improving the overall availability. The availability of the railway will also be improved through a reduction in human error.

4.3 Service benefits

The improvements in service benefits to the metro system customer that can be enabled through driverless operation will generally be through the improved visibility of staff in customer facing roles. The evaluation of these benefits will tend to vary between metros but broadly speaking the presence of staff will improve the safety and security of the metro through a reduction in anti-social behaviour. It will also enhance revenue protection and enable an increase in



accessibility to staff to provide passenger information especially to the infrequent users of the metro system. Driverless operation will change the working functions of operational staff, leading to more varied job roles and increased flexibility in staff deployment to fulfil these roles and benefits.

4.4 Summary of benefits

The benefits to the metro operator over and above conventional operation (including ATO) that may be realised through the implementation of driverless operation are briefly summarised in table 3.

Table 3: Summary of benefits.

Benefit	Mechanism
Improved journey time	• Reduced service reversing times • Consistent dwell times • Increase capacity • Staff independence from train operations • Removal of excess time due to driver actions (e.g. staff absences, breaks)
Increase operational flexibility	• Decouple crew plan from train schedule • Improve disruption recovery • Ease of introduction of new time tables • Multi skilled / flexible staff roles • Deploy resources in accordance with business / customer service needs
Reduced operating costs	• Reduce staff numbers • Reduction in safety duties of operations staff
Increase safety and security	• Reduction in Human error • Improved PTI protection • Increased staff visibility / presence

5 Systems engineering

5.1 System concepts

Having established the key drivers and reasons to automate the system, the next step is to capture the key concepts that characterise the system. These will be used throughout the development to ensure compatibility and achievement of the overall project goals. These concepts embody the reasons for undertaking the project and form the basis of the subsequent systematic analysis of the requirements. This document, next to the business case itself, forms the singularly most important document in the feasibility case. In developing this document the aim is to capture the operational vision and key decisions which will form the basis of the new driverless system. Subsequent analysis through a well formed systems engineering process, will subsequently refine these



concepts, but the important goal is to establish the fundamental changes and improvements that the enterprise wishes to make. The concepts of the new system and interactions of systems and people need to be assessed. The whole range of railway operational procedures and activities need not be covered in detail but merely in principle.

When considering the operating concepts it is important to bear in mind that conventional train operation is viewed as a system of operation consisting of people, processes and equipment, which operate within a particular environment to defined performance targets. The introduction of driverless operation seeks to enhance this capability with the introduction of new assets and systems. This increased capability enables the development of a system which allows for more effective operation of the railway as a whole. Driverless operation enables trains to be operated to the optimum levels of performance necessary to meet customer demand.

5.2 Migration concepts

When considering the feasibility of converting an existing operation attention needs to be given to the migration from the existing system to the new system. This also needs to include the operational organization. The operational concepts in each migration stage need to be understood and ultimately included within the system design. A number of options exist depending on the particular organisations starting point and experience with automation. The key migration principles need to consider the extent of the change being made and the experience of the operator to operate the railway in its new state (be it the final state or an interim migration stage). With a system operating as a 'line of sight', 'trip-cock' railway, the first migration stage would likely be to a 'Conventional' ATO railway, followed by migration to driverless operation with and then without an attendant on board, once the system reliability and operator confidence had been established. The level of automation that is achieved in each stage will depend on the satisfaction of both these factors, with the ability to return to a more restrictive operation should the need arise.

5.3 Systems integration

Clearly when automating an existing system, integration of the new system with existing assets needs to be understood. The various impacts that the proposed system will have, in the areas of operations, organisational changes, assets and advanced technologies, will need to be assessed. Trade-off studies should be used to assess the benefits of automation, for example in the areas of mixed mode operation a trade-off needs to be made between manual and automatic operation of the system. The arrangements for systems integration will need to be planned for all subsequent phases of the project. A key element of this at the outset is the development of the specifications that will be used to design the driverless system.

The requirements analysis process should completely specify the requirements of the driverless system that will satisfy the operational and



business requirements, developed through a systematic engineering process as outlined in Figure 2. A first step in this process is the capture of the business requirements to establish the overall objectives for the automation of the system. This should involve sufficient stakeholder consultation to ensure that the objectives are correctly captured. An operational model should be developed in conjunction with the metro operator to characterise the required behaviour of the driverless network through the development of operational scenarios. It should identify the functions required to automate the train operation, through consideration of new, existing and special operational conditions such as mixed mode operations required by the metro operator. The system boundary should be used to define those systems that will be changed as a result of driverless operation. Safety, RAM and Performance models should be used to develop targets and requirements. These models and resulting analysis can then be used to refine the business requirements into a set of system requirements for the driverless system.

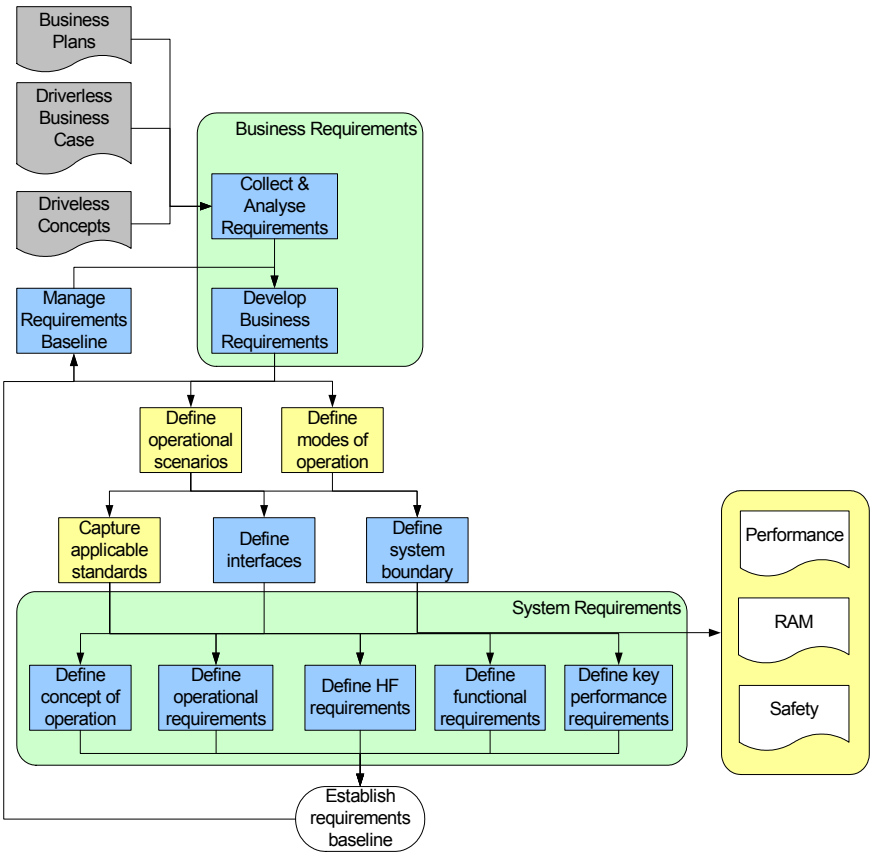


Figure 2: Requirements analysis process.



6 The business case

Having established the system concepts, operational reasons, additional safety measures, expected benefits and disbenefits, the business case to automate the railway can be produced. This will need to provide an economic evaluation of the proposed system using the parameters relevant to the particular enterprise and will need to consider the operational model to be employed. Clearly the evaluation needs to assess the lifecycle costs of automation. The business case can be developed on a progressive basis throughout the feasibility phase of the project to establish an initial case and final case, as more detailed specifications and costs are developed.

7 Conclusions

The feasibility of automating an existing heavy metro system requires consideration of a number of complex factors. The most significant of these is the constraints with the existing system, the improvements required and the benefits that will be delivered through these changes. To realize these benefits a holistic approach to the feasibility, specification and implementation of the system needs to be undertaken and needs to include the operational organization. Driverless operation is an enabling technology for alternative models of train and station staff deployment requiring both technological and organisational changes. Through understanding of these factors the feasibility case can be assessed and significant improvements to the journey time, service benefits and operational costs can be achieved.

Reference

- [1] Bill Gates, <http://www.quotedb.com/quotes/3854>.



Converting existing service to fully automatic operation

F. Fabbian

MTR Corporation Limited, Hong Kong

Abstract

Fully Automatic Operation (FAO) is rapidly becoming the system of choice for Mass Rapid Transit Systems. The technology is mature and there is now over 20 years experience in this technology. Existing systems would also benefit by being automated. The challenges of converting an existing manned system to FAO are more challenging than a green field installation. The MTR Corporation of Hong Kong is currently operating one line which is FAO and has a plan to convert the remaining lines. This is a system that transports 2.6M per day and trains that carry up to 2,700 passengers. This paper will share the planning strategy and business case for this conversion – the Hong Kong story.

Keywords: fully automatic operation.

1 Introduction

This paper will describe the automatic operation developments undertaken in MTRCL including the automatic turnaround function at terminus, automatic train door opening and closing, fully automatic operation for the new Disneyland Resort Line and the studies undertaken for the conversion of the existing operation to fully automatic operation.

MTRCL has adopted a progressive approach to the development of fully automatic operation. The automatic turnaround function at terminal stations was introduced in 1999 which was followed by the implementation of automatic train door opening and closing functions for the existing fleet and the fully automatic line, Disneyland Resort Line.

Feasibility studies were conducted to review the possible technical and operational impact and opportunities for converting the existing operation into fully automatic operation. These included a review of the operational



requirements, development of functional requirements for fully automated operation, reliability and risk assessments, external peer review of the technical specifications, and an energy study on potential savings after converting to fully automatic operation.

2 Automatic turnaround

Before the modification for automatic turnaround operation in 1999, the turnaround operation at a terminus was manned by two train operators, one at each end of the train. This was to allow the train to be driven to the designated turnaround track by one operator at the front or arrival cab who would then shutdown the cab so that the rear or departure cab at the other end could be started by another train operator for its return journey

To support the automatic turnaround function, both the trackside and trainborne ATC equipment were modified to support inter-cab communication through trackside telemetry and automatic train control authority transfer from the arrival cab to the departure cab thereby replacing the manual operation. With the modifications, the whole turnaround operation would simply be initiated by the train operator at the arrival cab. The train operator would then walk to the other end of the train to continue the journey in the opposite direction. Thus, the total manpower requirement was reduced.

In addition to staff cost saving, the automatic turnaround project provided the additional benefit of service consistency of train operation. It also provided the opportunity for MTR to review the operation requirements, method of train failure handling and operational rules and procedures when the system is under automatic turnaround operation. It was identified that new operational procedures were required to handle the following scenarios:-

- (a) Broken rail within the turnaround areas: the operational rules would now require the first trip at the beginning of daily operation must be manned by operators to visually check the rail condition;
- (b) Passenger trespassing at terminal station with no platform screen door (PSD): the operational rules would now require for non-PSD terminal station that the departure driving cab must be manned when the train is entering the terminal station platform.

3 Automatic train door opening and closing

A new circuit was built to interface with train door control circuits to enable train doors to open automatically when trains stopped at platforms. The new control circuit was built by programmable logic control (PLC) which would check the traction control circuits, zero speed detection relays and mode selector position to ensure train doors would only open at the correct side when a train was stopped correctly at the platform stopping point. A dwell time saving of 1.5s was achieved with this new function. A service trial was conducted with all trains converted in early 2005.



The new control circuit was subsequently expanded with the additional interface of digital voice PA to achieve the automatic train door closing function. A lookup table of preset dwell time of each platform was stored in the PLC. The PLC would start counting down the preset dwell time when a train is stopped at a platform. The digital voice PA would be triggered when the count down of 10 s is reached. The automatic train doors closing command would be issued when the preset dwell time is reached. A service trial is underway with 13 trains converted to monitor any service related impact.

4 Disneyland Resort Line

MTRCL awarded a signalling contract in 2002 for the seventh operational line – Disneyland Resort Line (DRL) for the Hong Kong Disney Theme Park. This new line which was opened in August 2005 was designed for fully automated operation with two stations and two 4-car trains.

The signalling system of DRL employs the ALCATEL's SelTrac40 Communication Based Train Control (CBTC) solution, which was designed for automatic operation. The ATP system of SelTrac40 ensures safety by maintaining the minimum train separation of one fixed virtual block. All block lengths were calculated in accordance with civil engineering speed limitations and train characteristics. The ATP system also provides the following features to safeguard the train movement:-

- (a) Rollback protection – The ATP safeguards the train movement in the assigned direction of travel by detecting the route set, operation mode and other trainborne inputs. If a backward movement over 1m is detected by the ATP, emergency brake is activated.
- (b) Overspeed protection – The ATP limits the speed of trains according to safe operating and civil engineering speed limits in all driving modes. The Emergency brake will be activated by the ATP whenever the train speed is above the ATP limit.
- (c) Door interlocking – The ATP interlocks with the train door control circuit to ensure the correct side of train doors can only be opened at the designated position in the platform when the train speed is zero and the service brake is applied.

The ATO system of DRL regulates the train speed using a combination of feedback and open-loop control algorithms to achieve and maintain the target speed imposed by the ATP system and supported train movement within the passenger ride quality criteria, i.e. maximum jerk limit of 0.2m/s^3 for all inter-station runs. The ATO system provides for the programmed station stops including train door opening and closing according to the station dwell time control as well as on-board audio announcement control and cab display. The ATO also supports the following non-vital functions:-



- (a) Remote control from ATS – To support driverless operation, ATO interfaces with other trainborne sub-systems to respond upon receiving remote commands from ATS. These commands included emergency brake reset, train hold, train door open/close command, service speed adjustment, VOB switchover, pantograph control, reset to rolling stock traction or auxiliary power circuits, etc.
- (b) Status Reporting to ATS – ATO makes use of the wireless LAN to report to the ATS continuously the current train status, e.g. train ID, operation mode, emergency brake status, train speed, door status, train hold status, travel direction, station stopping status and rolling stock alarms.
- (c) Energy saving in peak and non-peak hours – ATO adjusts train speed according to the ATS regulation strategy based on the energy saving strategy for peak and non-peak hours. By adjusting the dwell time and travel time under different regulation strategies, the train can be driven at the optimum speed to minimise the energy consumption at different periods of a day.
- (d) Audio announcement control – ATO automatically generates announcement requests to trigger audio messages at pre-defined train locations along the path of the train.

Without track circuits, the train detection was achieved by wayside tag, trainborne tachometer updating and continuous communication between On-Board Control Unit (OBCU) and Wayside Control Unit (WCU) through a wireless network. To boost the system reliability, the CBTC system made use of the dual redundant OBCU's from both ends of the train to support the features of rear-end automatic operation and remote speed restriction movement, which would allow a train to move forward to retain localization when the OBCU's at both ends lost the localization information.

To support maintenance works by locomotives within the new line, the system was customised to allow non-ALCATEL equipped trains to move safely within DRL and interface with ALSTOM's signalling system in Tung Chung Line (TCL) to ensure a smooth and safe cross line operation. This was achieved with the introduction of axle counters at fixed locations along DRL thereby creating a traditional fixed block operation for route setting when a non-ALCATEL train is running inside DRL area.

The ALCATEL wayside equipment was extended and overlaid on the existing ALSTOM wayside area so that the DRL train could be operated between the DRL area and the depot in Siu Ho Wan of TCL. The new signalling system was designed to protect DRL trains and provide all the ATP functions in Coded Manual operation mode using track occupancy, point position and signal status information from the ALSTOM system.

This inter-operability design experience provided useful invaluable inputs in overcoming similar challenges of conversion in future.



5 Studies for fully automatic operation

In April, 2004, a working group was formed to look into the different aspects of converting the existing train fleet (106 trains of 8-car consists) to support fully automatic operation.

The benefits identified for adopting automatic operation included improved regulation of train service, saving in manpower and energy. Moreover, automatic operation would enable the train service regulation to be more responsive to customer demand especially during festive occasions and peak hours. This improved responsiveness would also accelerate service recovery after perturbations. Automatic operation would also improve system reliability by removing variable human errors, mostly caused by monotonous work. Experience from other railway operations indicated train service could be more reliable than other lines with train operators who would be redeployed to other value added tasks of better customer service.

The working group analyzed the existing railway systems, including rolling stock, signalling, communication and other infrastructure, to determine the gaps for a fully automated operation railway. To support a fully automated operation railway, our systems must be retrofitted to cope with the following functional requirements.

- (a) Automatic train door control to support platform duty according to pre-defined platform dwell
- (b) Automatic train starting when ATP allows the train to depart after a series of automated checks and confirmation sequence are performed.
- (c) Automatic jogging movement when a train stops within a limited distance away from the designated stopping point.
- (d) Two-way communication between Operation Control Centre (OCC) and passengers inside saloon.
- (e) Remote monitoring of train status by operators in OCC so that operator can visualize the health status of the trainborne equipment remotely.
- (f) Remote control from OCC to trainborne equipment, e.g. fault reset to ATC & auxiliary equipment, pantograph control, train initialization checking, etc.
- (g) Upgrade the existing radio network to support the real-time two-way communication between trainborne equipment and OCC.

Train modification would be required on train door control and provisions of interfaces for automatic operation to support jogging movement, remote control and fault reset. A highly reliable radio system is required for maintaining communication to passengers during normal and emergency situations and the radio system is essential for remote control and fault resetting functions.

To mitigate platform edge hazards such as passengers falling on track, platform screen doors or intrusion detection systems would be required at all stations. The market suppliers were invited to comment on the proposed requirements and the system that they may offer.



6 Reduced power consumption

Apart from the long-term saving of the re-deployment of train operator staff, another major saving is traction energy. A study to quantify the possible energy saving of automatic operation was carried out.

The study included an in-depth review of our system configuration including the track layout scheme, electrical infrastructure and the characteristics of the train fleet, and service pattern (e.g. timetable and passenger activities on platform). A traction energy consumption model of the existing system configuration was built. Based on this model, the expected energy saving was simulated for different operation scenarios.

The simulated result concluded that since automatic operation could effectively minimize the turnaround time at terminals, the saved time could be used to extend the inter-station run time and thereby reducing the energy consumption. Extended run time meant that trains could be run at a lower speed to save energy. Another major contribution of the shortened turnaround time was the reduction of the number of trains required for the same service requirement. Trains may also be out-stabled at sidings hence avoiding the energy cost of transferring them to and from the depot.

7 Service performance target

One of the key objectives of the working group is to define the appropriate service performance targets of automatic operation.

Instead of using the performance target from other railways which may not be fully applicable to MTRCL, we have adopted a bottom-up approach in projecting the expected performance based on the existing system performance with the expected new functions.

For instance, there is currently little service delay arising from a minor overrun or stop short at a platform stop since the train operator can respond immediately to resolve the situation. However, under automatic operation, this incident will probably cause a service delay if no automated corrective measure was introduced. To address this issue, a new functional requirement of automatic train jogging is required when the system detects a minor overrun or stop short.

After completing this exercise to review all incident scenarios, it was concluded that improved system reliability must be achieved in order to maintain the same current service performance. This review also assists in defining the necessary requirements of automatic operation as well as the necessary operational procedures.

8 Human resource and community issues

A major challenge with implementing fully automatic operation on an existing railway is to have the staff appreciate the company's vision for service enhancement and ambition to meet customer expectation. A well planned HR strategy and communications will have to account for the affected



staff numbers, run-down by natural attrition i.e. retirement or departure, and staff deployment to new extensions. There will also be the opportunity to multi-skill and redevelop skills of some staff to handle other responsibilities in the railway system. For example, more staff can be put on the network to enable a more flexible and responsive system to meet the needs of customers when they are travelling in the system.

The successful application of fully automatic operation in the Hong Kong Airport People Mover, and recently the MTRCL's Disneyland Resort Line has rendered fully automatic operation gain wide acceptance by legislators, politicians, public and MTRCL's own staff. More extensive implementation in the existing railway lines will be complemented by a PR plan to manage community issues e.g. safety, loss of jobs. The PR plan will cover study visits to other railways running on fully automatic operation for the key players such as legislators, regulators and politicians.

9 Conclusion

MTRCL has taken a progressive approach to the introduction of automatic operation, i.e. automatic turnaround, automatic train door opening / closing and Disneyland Resort Line which is fully automatic.

Based on the experience obtained, it is concluded that the technology for automatic operation is mature with a proven track record. This technology would provide a good means to match the service supply to demand both economically and commercially.



This page intentionally left blank

Re-signaling the Paris Line 1: from driver-based to driverless operation

C. Braban¹ & P. Charon²

¹*Siemens Transportation Systems, France*

²*RATP, Paris Urban Transport Operator, France*

Abstract

In November 2005, RATP – Paris Urban Transport Operator – awarded Siemens a contract to upgrade the oldest line of the Paris metro to driverless operation (with no driver onboard). This paper discusses RATP's motivations and focuses on the technical challenges for upgrading the ATC from driver-based to driverless operation. This paper also presents the Communications-Based Train Control solution engineered by Siemens.

Keywords: communications-based train control, re-signaling, driverless Automatic Train Control.

1 Introduction

Since the beginning of the 2000s, RATP, Paris Urban Transport Operator, has undertaken a vast re-signaling program to modernize the automatic train control (ATC) systems installed on its metro network. The overall program aims to increase safety and transport capacity, and improve passenger comfort.

The OURAGAN re-signaling program, focusing on driver-based train operation (Paris lines 3, 5, 9, 10 and 12) to which Siemens Transportation Systems already contributes is part of an overall modernization scheme. In addition to improving safety, capacity and passenger comfort, it aims to achieve high level of parts interchangeability in the communications-based train control system to insure its easy maintainability.

In conjunction with OURAGAN program, feasibility studies have been conducted by RATP to upgrade the existing lines with strong transport demand to driverless lines, following the example of line 14 which is in revenue service since 1998 and driverless from day one.



The automation of line 1 is achieving two goals: the upgrading the PA BF (speed code type of automatic train control solution engineered by Siemens in the early 1970s) to Communications Based Train Control (CBTC), and secondly conversion to full driverless train operation (with no driver onboard), the success on the line 14 is now widely recognized.

The paper explains the motivations of RATP and focuses on the technical challenges for upgrading an existing driver attended line to a driverless one. This paper also presents the Communications-Based Train Control solution engineered by Siemens.

2 The stakes of upgrading Line 1 to driverless train operation

2.1 The oldest line but also the most loaded

Opened to the public in 1900, Line 1 extends across Paris from the east to the west, a distance of over 17 km. It is the oldest metro line, and also the most prestigious metro line because its 25 stations serve most of the places of interest in the French capital, as well as numerous business districts and shopping centers.

The annual passenger traffic on Line 1 is, at 207 million passengers, the highest in the whole metro network, making it the most heavily used line of the Paris metro. One of the reasons for the high ridership lies in the line's strategic layout, serving 16 of the most crowded stations, as well as 5 out of 15 major multimodal nodes of the network.

A fleet of 52 MP89 trains operate daily, with a headway of 105 seconds. 23,500 passengers are transported per hour and per direction. Thanks to a commercial speed of 27 km/h, the estimated travel time from Château de Vincennes (in the east) and la Défense (in the west) is only 35 minutes.

The characteristics of line 1 mentioned above lead to a high request for adaptability of the transport offer, which is hardly reachable with a driver operated line.

Another factor in support of the choice of fully automating line 1 is the necessity of renewing the current equipment before 2010. This particularly concerns signaling, OCC, ATC and interlocking equipment which are 4 major parts of the automation system. Indeed, the current signaling equipment needed to operate the line consists of:

- The Operation Control Center, which was installed in Boulevard Bourdon in Paris, in 1967 and upgraded in 1981.
- The wayside signaling used for safe train separation, which was installed in 1956.
- The PA BF (Pilotage Automatique Basse Fréquence), the speed code solution engineered by Siemens, used for controlling train movement.

The changeover to driverless train operation will increase transport capacity thanks to a time headway reduction from 105 to 85 seconds. The speed of driver-attended trains is currently limited to 70 km/h. With driverless trains, the



maximum speed rises to 80 km/h, enabling the commercial speed to be increased.

The increasing proportion of delays due to passengers also contributes to the choice of line 1 automation. 72% of the delays are due to passengers and among them 69% can be controlled by using platform screen doors (e.g. serious accidents, alarm signal, passenger on tracks, objects on tracks).

Moreover, from an economical point of view, line 1 automation will reduce the operating costs between 10 and 15%, without taking into account the accidents that are avoided doing so.

Finally, the positive feedback of satisfaction from both passengers and operators on line 14 and the above-mentioned elements convinced RATP that line 1 automation is the best solution.

The line 1 automation program consists of different kinds of projects. Most of them obviously deal with technical issues such as the automatic train control system, which is the main subject of this paper, and also the platform screen doors, the rolling stock, the audio-visual equipment. But there is also an organizational project to upgrade the line operation and jobs. As a consequence, this ambitious program, whose objectives and constraints are shared by many transport operators and authorities around the world, requires a complex financial, social and technical process of design and management under the responsibility of RATP. The main milestones of this program are:

- the start of the platform screen doors installation in January 2008
- the commissioning of the new Operational Control Centre in February 2009
- the first operating driverless train in October 2009
- a fully automated line in 2010

2.2 The CBTC Line 1 upgrade contract

On November 7, 2005, RATP and Siemens Transportation Systems France signed a 30.8 million euro contract for upgrading Line 1 from driver-attended to driverless train operation, thus consolidating Siemens' leadership in the design of automatic train control (ATC) solutions for the RATP. This new award follows a number of ATC solutions designed for RATP by Siemens:

- Automatic Train Control for driver-attended lines: PA 135 and PA BF installed on 12 of 13 metro lines.
- Automatic Train Control (SACEM) for RER A and B regional Express lines.
- Driverless Communications-Based Train Control for Line 14 and its extensions to Saint Lazare railway station (in revenue service since December 2003) and Olympiades (in revenue service in 2007),
- Communications-Based Train Control for driver-attended lines as part of the OURAGAN re-signaling program: lines 3 and 5, and subsequently lines 9, 10 and 12.



The upgrading Line 1 to a driverless line presents real technical challenges. The challenges are unique because the switchover from the current operation (using a fleet of trains with drivers under the control of a speed code solution) to driverless operations (using a free-propagation CBTC solution) will have to be carried out without any disruption of passenger service.

The contract comprises the:

- Design and supply of equipment necessary for driverless train operation: wayside equipment and onboard equipment installed on 49 cars.
- Supply of the Data Communication System based on free propagation radio, comprising the wayside backbone network and equipment for wayside-to-train communication.
- The Operation Control Center (OCC), which enables RATP to supervise and manage operations on the line.

The cut-over strategy that enables the switchover from driver-attended train operation to driverless operation consists of three key milestones:

1. Refurbishment of the existing OCC to meet specific requirements of future driverless train operation. The new OCC is then put in service to supervise and manage current driver-attended trains operation.
2. Operation of the first driverless train.
3. Mixed-mode operation: both manual operated and driverless trains run on the line during a interim period. Progressive replacement of driver-operated trains by driverless ones until the entire Line 1 fleet is running in driverless mode.

To achieve these milestones a number of technical requirements must be satisfied:

- The system will have to support mixed-mode operation: During a given time period, driver-operated and driverless trains will share the line during daily operation.
- The installation and testing of the system will have to be performed as transparently as possible with respect to daily passenger service. As a consequence, the only possible interruptions of passenger service have to match the normal daily service interruptions, namely at night during an extremely short three hour time window.

3 The CBTC solution designed for Paris Line 1

3.1 Key assets

The technical solution designed for Line 1 allows for:

- Simultaneous and safe operation of driver-attended and driverless trains thanks to the tracking of CBTC-equipped and unequipped trains.



- Unequipped trains, operated by drivers, are tracked using the existing track circuits. Movement authorities are delivered to drivers via the existing wayside signaling equipment.
- Driverless trains are tracked based on the location report issued by onboard equipment and delivered to wayside equipment. Movement authorities are then communicated to the trains by wayside equipment based on an exchange of messages. This is the fundamental principle of Communication-Based Train Control.
- Minimal installation work thanks to a significant reduction of wayside equipment as the result of:
 - The use of the free-propagation radio solution engineered by Siemens, adaptable to free ISM band frequencies as well as to narrow proprietary bands. The wayside equipment is installed in stations.
 - The decentralized system architecture, with equipment installed in technical rooms located in stations, thus reducing the cabling along the line.

To reduce the number of tests performed during daily operation and to maximize the time available for testing at night, Siemens devised a system qualification strategy, based on:

- The design of a simulation platform allowing the entire system to be tested in the factory.
- Tests performed during daily operation in so-called “shadow mode”: this permits the observation and analysis of the behavior of the system installed in its real-life environment over a long-time period without impact on passenger service.
- The implementation of sophisticated migration processes, making it possible to switch from the current mode to driverless operation and back again within the time slots available at night dedicated to testing.

3.2 Focus on train tracking

The simultaneous, safe operation of driver-attended and driverless trains relies on tracking system based on virtual blocks combined with the moving block principle. This stems from the technical evolution of tracking systems currently in use on the Meteor Line, on the Canarsie Line, on Barcelona line 9 and on OURAGAN.

For the tracking of unequipped trains, the existing track circuits are used. The track is physically divided into “blocks”, which only contain one train at a given time. Entry to each block is protected by a signal light (red/green) that informs the driver whether or not the block is already occupied by a train.

The tracking of CBTC equipped trains does not depend on track circuits. Instead, the track is split into “virtual blocks”, which overlay the existing physical blocks. Each train calculates its own location on the line and transmits it to the wayside CBTC equipment.



The wayside CBTC equipment continuously updates the status (occupied or free) of each virtual block based on:

- The location reported by the onboard CBTC equipment and delivered to the CBTC wayside equipment, using free propagation radio.
- The occupancy status of track circuits.

The wayside CBTC equipment then computes for each train a target point, i.e. a danger point, not to be passed in order to prevent a hazardous situation from occurring. The target corresponds to the first danger point located downstream of a train. It can be a block (a track circuit or a virtual block) occupied by a train, a red signal or the end of a territory under CBTC control.

Using free-propagation, the wayside CBTC equipment delivers the target point to each CBTC equipped train. In the event that the target point corresponds to an occupied virtual block, the target point is set to the last reported position of the train ahead. This is the moving block principle. If the target point corresponds to a track circuit occupied by an unequipped train, the target point is set to the entry to the track circuit.

Equipped trains then compute the ATP (Automatic Train Protection) speed curve (which guarantees that the train will come to a stop before the next danger point) and the ATO (Automatic Train Operation) curve.

In the case of unequipped trains, information about the next danger point is delivered to drivers using the existing wayside signaling. A green light means that it is safe to enter the block as it is not occupied by any train. A red light means that it is forbidden to enter the block as it is occupied by a train.

3.3 Focus on the free propagation radio solution

Now in operation on the Canarsie Line in New York City, the free propagation radio product offers outstanding performance in the face of the stringent requirements CBTC in underground railway applications. Indeed, the challenge taken up by Siemens was to adapt the various technical innovations to the specific needs of underground transit applications by favoring system availability in these difficult environments for the propagation of radio waves.

For driverless train operation, availability of the data communication system is essential to enable the CBTC system to safely replace the driver. The radio solution fulfils the following requirements:

- Robustness to guarantee the very high level of availability required, despite the diversity of the environments: alternating between tunnels and open air sections, interference and masking. The signal modulation is based on Direct Spread Spectrum Sequence (DSSS). The demodulation takes into account the energy carried by the various multi-paths to “rebuild” the signal transmitted.
- High performance with respect to message updating and handover constraints. Wayside-to-train messages can be addressed to all the trains or to a set of trains on the line. Messages are refreshed frequently



(approximately every 0.5 s) to ensure maximum performance of the CBTC system.

- Robust intrusion protection to ensure security. As the identity of the communication is determined, there is no need for standardization at the air gap. No interface at the air gap is a guarantee of protection against intentional or accidental intrusion.
- Proven coexistence with ISM band users: no perturbation from WIFI users, no perturbation to WIFI users.

The robustness of the free-propagation radio system was achieved through the thoroughness with which the radio link budget was drawn up in various environments as well as: TDMA and micro-synchronization, diversity, and geographical organization redundancy. As a consequence, it guarantees unrivalled operational availability for a reduced amount of wayside radio equipment. A total of 55 radio equipment installed on 17 km of the Canarsie Line.

For Line 1, this feature is of key importance as it contributes to the simplification and minimization of the installation work since the free-propagation radio system does not require any continuous medium on the guideway and wayside radio equipment is strategically located in stations.

3.4 Simulation platform

In order to minimize the number of tests performed on the line during revenue service hours, and so reduce the risk of disruption to passenger service, considerable efforts are dedicated to the design of powerful simulation tools.

These tools allow extensive in-factory testing including the functioning test of an item (which can be software, a piece of equipment, a subsystem or the complete system) and evaluation of its performance in an environment closely resembling that on the line. The tools are based on key modules comprising the OCC, the wayside signaling, wayside CBTC equipment, CBTC equipped trains, platform screen doors and free-propagation radio system and their corresponding interfaces. Behavior of the items will be assessed both in nominal and degraded conditions based on predefined scenarios. It is possible to interact with the simulation at any time in order to generate events or create failures, thereby increasing the relevance of the test performed.

3.5 The migration strategy

The major challenge when re-signaling lines in revenue service lies in the validation of the system – in the specific case of Line 1, the OCC and the CBTC onboard equipment – without impairing the quality of service offered to passengers during daily operation and compromising safety. A fundamental feature of this solution is therefore the ability to be operated in “shadow mode”, whereby the system receives and sends all necessary information as if under CBTC control, but without any actual outputs being activated.



A second challenge which makes this project unique is the coexistence of:

- Totally different automatic train control technologies – speed code and CBTC,
- Different modes of train operation – manually operated trains and driverless trains.

The migration strategy is organized into three main time periods corresponding to the three key milestones scheduled by RATP.

- *Period 1* is concerned with all the installation and testing work requested for operating the line with the new OCC. The main technical difficulty lies in testing and commissioning all the functions available in the “new” OCC, while still ensuring the safety of daily train operations under the control of the “old” OCC. This work is carried out during both daily operations and night-time hours. It involves having the two OCC working in parallel: until the first milestone is reached, the “old” one remains active and the “new” one operates in shadow mode; at the end of this period, the “new” OCC becomes active, but the “old” one remains available for operation if necessary.
- *Period 2* focuses on the testing and commissioning of the driverless CBTC solution to enable the first driverless train to operate on the line together with manually operated trains. It should be noted that all the wayside CBTC equipment is installed during the Period 1. The testing covers ATP and ATO functions. The main difficulty is how to perform all the necessary tests and customizations without interfering with work already done. For this, the system will first be extensively tested on the test track in Valenciennes, France. In addition, a CBTC system installed on a MP 89 train will be tested in shadow mode on Line 1.
- *Period 3* is concerned with the progressive introduction into revenue service of the new MP05 rolling stock equipped with the driverless CBTC. Manually operated trains are progressively removed from service. The final functionality of the system is tested, including automatic train regulation and management of the depot.

4 Conclusion

The upgrading of Line 1 of the Paris metro from a driver-based to a driverless operation (with no driver onboard) is undoubtedly one of the most important projects RATP is going to realize in the next five years.

After the great success achieved with Meteor Line 14, which demonstrated the undeniable advantages of driverless operation, this project opens up the way to re-signaling and upgrading conventional lines to driverless operation.

Two major stakes are associated with the line 1 project. The first one addresses the unquestionable improvement of the quality of service, resulting from the flexibility of driverless operation. The second one is related to complex political and social issues associated with it.



Switching over from a conventional line operation to a driverless mode of operation is not a simple matter. It includes major technical challenges, i.e. to install, test and commission the complete system in total transparency to daily operation, so as not to inconvenience passengers or reduce safety in any way.

Since its first implementation on Line 14, the solution proposed by Siemens has benefited from the latest technical innovations implemented on the Canarsie Line of New York City Transit, Line 9 of Barcelona metro and lines associated with the OURAGAN program. It not only makes the most of its key features such as the tracking of both equipped and unequipped trains, the free-propagation radio, the system's ability to operate the system in "shadow mode", but also integrates state-of-the art simulation tools allowing the testing in real-life scenarios without disrupting operation to successfully take up the technical challenges.

As for RATP, the success of Line 1 Automation relies on the skills and the technical control of the selected suppliers as well as on the experience and the expertise of the RATP in the operation control, the project management and the system risks management.



This page intentionally left blank

Application of communication based Moving Block systems on existing metro lines

L. Lindqvist¹ & R. Jadhav²

¹*Centre of Excellence, Bombardier Transportation, Spain*

²*Sales, Bombardier Transportation, USA*

Abstract

The unique features of Communication Based Train Control (CBTC) systems with Moving Block (MB) capability makes them uniquely suited for application 'on top' of existing Mass Transit or Metro systems, permitting a capacity increase in these systems. This paper defines and describes the features of modern CBTC Moving Block systems such as the *Bombardier* CITYFLO* 450* or *CITYFLO 650* solutions that make them suited for 'overlay' application 'on top' of the existing systems and gives an example of such an application in a main European Metro. Note: *Trademark (s) of Bombardier Inc. or its subsidiaries.

Keywords: CBTC, Moving Block, CITYFLO, TRS, Movement Authority, norming point, headway.

1 Introduction

The use of radio as a method of communication between the train and wayside in Mass Transit systems, instead of the traditional track circuits/axle counters and loops is gaining popularity. The radio based CBTC systems are uniquely suited for application 'on top' of existing Mass Transit or Metro systems for increased traffic capacity as CBTC systems normally do not interfere with the existing systems. This allows an installation of the CBTC system in a line in operation whilst maintaining full safety and capacity during the process.

The fact that CBTC systems also allow Moving Block operation adds to the possible increased traffic capacity that can be achieved with such systems.



2 Definitions

2.1 Communication Based Train Control (CBTC)

Although the term Communication Based Train Control in theory allows for any ‘contact-less’ communication between train and wayside, in this paper the term is used to designate the more modern type of CBTC system using radio as the communication medium.

2.2 Moving Block (MB)

The traditional Mass Transit systems using track circuits or axle counters as a method for detecting the presence of the train are ‘fixed block’; the block being defined as the fixed length of the track circuits and axle counters. In CBTC systems radio is used as the communication medium, enabling the position of the train to be sent by the train itself and in turn making it possible to have a ‘moving block’ operation (or more accurately a ‘moving and variable block’ operation) as there is no equipment with fixed lengths in the system.

A moving block system allows the trains to run closer to each other compared to a conventional fixed block system, thus reducing the possible headway.

However, CBTC systems may also operate within a ‘fixed block’ mode, if so desired, thus permitting increased compatibility with traditional systems, while compromising on the headway.

2.3 Movement Authority (MA)

A Movement Authority is defined as the authority for a train to safely proceed up to a certain point where it has to stop. In fixed block systems, the Movement Authority consists of a locked train route starting at a certain signal with a proceed aspect and ending at another signal with a stop aspect, passing through one or more track sections.

In a CBTC *fixed block* system the Movement Authority is set from the predetermined block point where the train is to a predetermined point on the track, normally the end of a track circuit or similar.

In a CBTC *moving block* system the Movement Authority is set from the exact point where the train is to a ‘conflict point’ in the track ahead of the train.

In a CBTC system with constant update of information about the train’s position and constant renewal of the Movement Authority the train will be allowed to proceed without braking as long as there is no conflict point within braking distance ahead of it.

2.4 Conflict Points (CP)

A Movement Authority for a train always ends at a ‘conflict point’ ahead of the train. A ‘conflict point’ is defined as:

A location along the track beyond which a train *NOT* permitted.



A CBTC system utilizes these conflict points to properly and safely manage the movement of trains throughout any metro line. A conflict point can either be *static*, meaning that its location in the track is fixed or *dynamic*, which means that its location is a moving train. An example of a static conflict point is a buffer stop at the end of the line and an example of a dynamic conflict point is the end of the train in front.

Furthermore, a conflict point can have two states - *mutable* or *immutable*. A mutable conflict point can be either *active*, meaning it is a conflict point, or *inactive*, meaning it is not a conflict point. Immutable conflict points are always conflict points.

Examples of typical conflict points are:

Table 1: Typical conflict points.

Conflict Point	Type of Location	State	Active/Inactive
Rear of Train in Front	Dynamic	Immutable	always active
Buffer stop	Static	Immutable	always active
Point	Static	Mutable	active / inactive

3 Elements of a typical modern CBTC system

Broadly speaking, a modern CBTC system can be said to consist of four parts:

- the **Control Centre System** which controls the operation
- the **Wayside System** which receives train positions and issues Movement Authorities to assure the safe running of the trains
- the **Vehicle System** that generates the train position and receives the Movement Authorities and assures the compliance of the Movement Authorities and
- the **Communication System** which allows the transfer of messages to and from the train.

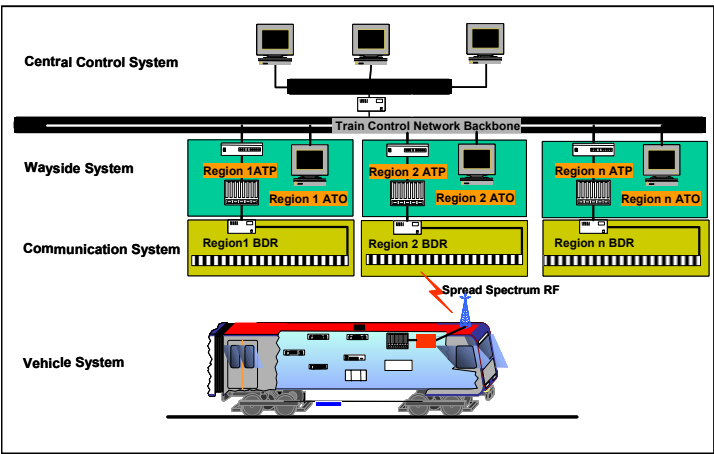


Figure 1: Block Diagram of a modern CBTC system.



3.1 The Control Centre System

The Control Centre System normally consists of ‘off the self’ servers and operator’s work stations that run the CBTC application. The Control Centre System allows operators to direct trains from one location to another, turn them around at the end stations, or in the middle of the line, and permits trains to leave and enter the depot(s).

All modern CBTC Control Centres also have programs for automatic driving of the trains without operator intervention, allowing either regulation by time table or by headway.

Furthermore a modern Central Control System also has control of auxiliary functions like Passenger Information Systems (PID), Telephone Systems and CCTV systems. In many cases the Central Control System also contains the SCADA systems for control of auxiliary systems such as traction, escalators and air conditioning.

The Central Control System is always duplicated in modern systems in order to achieve the availability needed.

3.2 The Wayside System

For Mass Transit lines of normal lengths, the wayside equipment is distributed along the line and divided into parts, often called *Regions*. Each region is responsible for safe movement of trains within its boundary of control and safe handover of the trains to adjacent regions.

The size of each region depends on the length of the line controlled by it and the maximum number of trains that need to be handled within the region.

The regions contain the ATP and ATO parts of the CBTC system responsible for issuing Movement Authorities and communicating train positions to the Central Control System.

The Wayside System components are operationally redundant for the highest availability.

3.3 The Vehicle System

3.3.1 ATP and ATO System

The Vehicle System mounted onboard the train contains the equipment needed to acquire all the information from the train and the track, to process it and to transmit the train position to the Wayside System. It also contains the equipment that receives the Movement Authorities from the Wayside equipment and displays this information to the driver and controls the driving of the train through the ATP and ATO.

The Vehicle System is designed in such a way that the systems installed in each end of the train can be used as replacement for each other thus creating a duplication of equipment, in order to achieve the highest availability.

3.3.2 Train Position System

Since the train in a CBTC system is transmitting its position to the Wayside System, the train must know where it is in the network. This is achieved using



the regions and then subdividing the regions into segments. Within the segments, the train is using its onboard tacho-generator and or other sensors to measure the distance from the start of the segment (offset). The figure below shows how regions, segments and offsets are used to define the train's position.

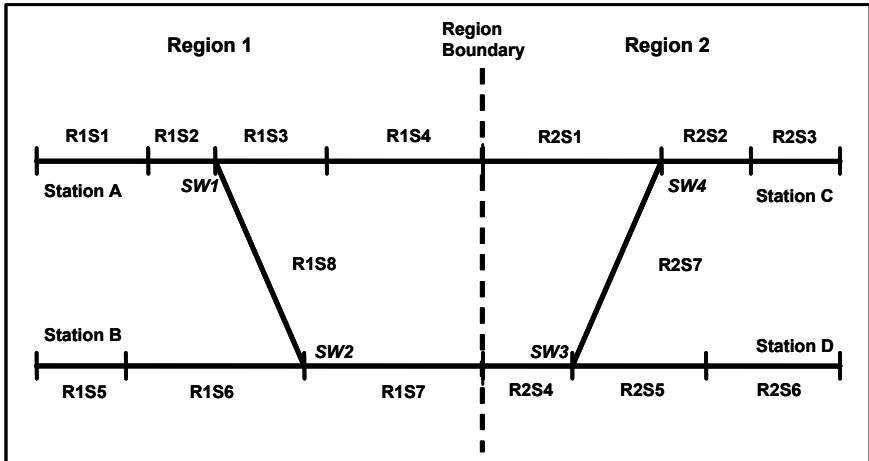


Figure 2: How regions and segments are used for train positioning.

In a *CITYFLO* 450 or *CITYFLO* 650 system, the train's position is sent to the Wayside System as region number, e.g. 'R1', segment number, e.g. 'S3' and offset from the start of the segment, e.g. 500m. The complete position would then be: 'R1, S3+500'.

To correct any errors in the position measurements made by the tacho-generator 'Norming Point' balises are used which are mounted along the track. When a train passes over such a balise it provides an exact location of the train and the Vehicle System can correct for any errors in the position of the train. As the train moves away from the 'Norming Point', the position error will start to increase and this will again be corrected at the next 'Norming Point'.

3.4 The Communication System

The communication system in modern CBTC systems is based on radio transmissions, often in the 2.4 GHz ISM band and often using a spread spectrum technique to reduce the chances of interference from other systems.

The communication medium can be based on 'line of sight, or leaky coax using a RADIAX cable, or both depending on the application.

All communication system components, except the antennas or the RADIAX cable, are also duplicated.

3.5 The Train Registry System

One of the few drawbacks of CBTC systems is at 'cold' start-up. When the system is started again after a total system power-down, the CBTC Wayside



System is unaware of the position of the trains. In earlier CBTC systems it was often necessary to have drivers board the trains and drive them manually to the next station in order for the CBTC system to 'acquire' the position of the trains in a safe way. Although system stops in CBTC systems are unlikely due to the fact that nearly all components are duplicated, having to drive the trains manually could be very time consuming, sometimes taking an hour or more for larger metro systems.

The *CITYFLO* 450/650 systems offer the Train Registry System (TRS) feature, which registers the identity of the trains as they pass in and out of the regions independent of the *regions*. In case of a 'cold' start or after a brief communication failure, the TRS system will provide the train IDs for each region to the CBTC Wayside System so that the communication can be re-established instantly thus making any start-up a matter of minutes.

4 Application example, overlaid CBTC system for European Metro

4.1 Background

Most European metros were established some time ago and by now many of them will need a modern signalling system for one or several of its lines in order to increase their transport capacity. Normally there is in such a metro an existing fixed block, speed-step signalling system, with an older ATP and ATO which is working at full capacity and can not be upgraded.

Many such metros recognize that to install a moving block CBTC system is the only way forward. The solution they often arrive at is to overlay a CBTC moving block system 'on top' of the existing signalling system in order to achieve a 'dynamic headway' of down to 40 seconds, i.e. a headway calculated without the station dwell times, and to be installed without affecting the passenger safety or transport capacity in the process.

As this often is their first experience with CBTC systems, most metros would like to have a conventional system as 'fall-back', should the CBTC system fail. Finally, although often the proposed CBTC system would have a driver onboard, the system itself is more often than not required to be capable of being upgraded to a fully driverless system.

The selection of the *CITYFLO* 450 system from Bombardier Transportation to re-signal a metro line with these requirements will lead to the following system solution.

4.2 The system solution

The requirements of overlaying the CBTC system on the existing signalling system along with providing a fall-back system, to assure full safety and transport capacity during the installation and to be able to meet the dynamic headway requirement of 40 seconds lead to a system solution shown below, and which uses several of the unique features of the *CITYFLO* 450 system.



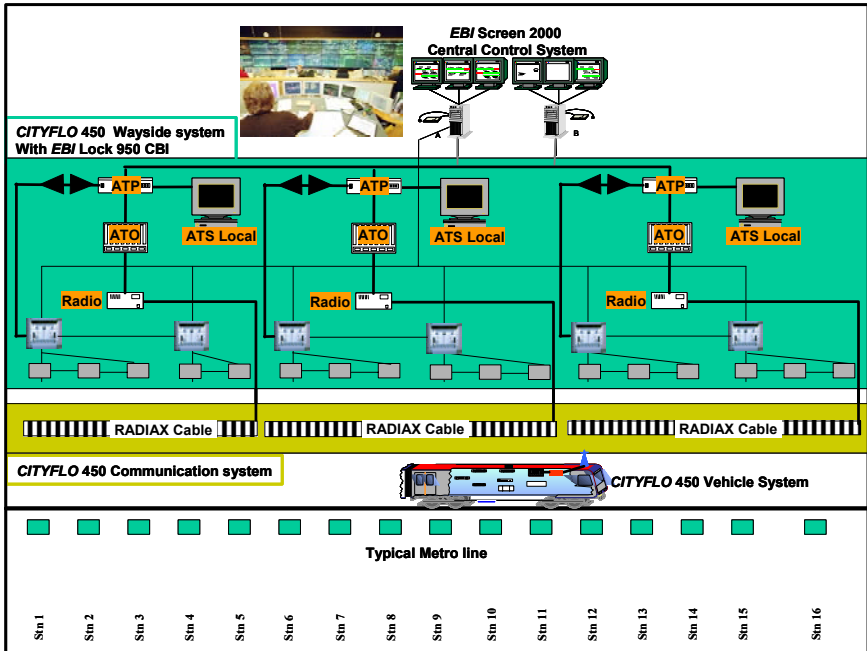


Figure 3: Typical Metro System solution.

4.2.1 Meeting the ‘overlay’ requirement

The *CITYFLO 450* system uses radio for communication between wayside and train. The train position is determined by the onboard ATC equipment and communicated to the wayside ATC over the radio. The system does not need track circuits or other form of wayside interface for safe operation and therefore, can be easily ‘overlaid’ on top of an existing signalling system.

4.2.2 Meeting the ‘fall-back’ requirement

CBTC is relatively new technology and even though metros around the world are beginning to embrace this new technology for the obvious benefits it is likely to bring, the approach is cautious. Choosing to operate the system with drivers initially is a result of such caution even though there are a number of metros which already operate driverless trains using CBTC systems. Requirement for a ‘back-up’ signalling system is another example of such cautious approach. Even though modern radio systems are highly reliable, due to the nature of the CBTC systems, a single loss in the communication chain can bring any system to a grinding halt. Even though highly unlikely, metros often require the system to continue to operate under such a situation.

In such cases Bombardier proposes a full secondary signalling system using the conventional track circuits, interlockings and wayside signals. The solution is based on the *Bombardier* EBI* Lock 950* Computer Based Interlocking (CBI) and TI21 jointless track circuits. *EBI Lock 950* CBI was first introduced in 1976 and is currently in its fourth generation. The interlocking has certain special



features that make it ideal for Mass Transit applications and especially for co-operation with the *CITYFLO* 450 system. Note: *Trademark of Bombardier Inc. or its subsidiaries.

In particular, as the interlocking allows the outputs to the wayside objects like signals, point machines and object controllers to be located at practically any distance from the central interlocking unit, it has been possible to use the capacity of each central unit in two to three metro stations using fibre optic cable as communication medium. This has allowed the number of interlockings needed for each line to be reduced.

The *EBI* Lock 950 CBI and the *CITYFLO* 450 systems communicate with each other over a safe serial link. While the *EBI* Lock 950 interfaces with the wayside objects, the *CITYFLO* 450 manages the communication between the trains and wayside ATC. The movement authority is generated as a safe and optimum balance between the trains' reported position and the actual track occupancies. In normal operation when the *CITYFLO* 450 system controls the operation of the trains, the computerized interlocking will act as little more than a conduit between the CBTC system and the object controllers controlling the wayside objects, with only basic functionality.

Should the *CITYFLO* 450 system fail, the affected trains will still be able to operate using the movement authority generated by the *EBI* Lock 950 interlocking and conveyed by means of the wayside signals.

For this back-up mode, it is possible to choose longer track sections in order to reduce the number of track circuits. Therefore the headway in fall-back mode will obviously be much higher than with the CBTC system.

4.2.3 Meeting the requirement to maintain safety and capacity during installation

The existing onboard ATP/ATO equipment in the trains will be removed when the *CITYFLO* 450 vehicle equipment is installed on a train. Therefore, it is necessary to choose a system solution that allows safe 'coexistence' between trains equipped with the old ATP/ATO and trains equipped with the *CITYFLO* 450 ATP/ATO without affecting the performance.

To this end, the *CITYFLO* 450 system will have information from the existing track circuits and will therefore distinguish between a train with the old ATP/ATO occupying a track circuit and a train with the *CITYFLO* 450 ATP/ATO occupying a track circuit. The latter train is also sending information via radio about its position while the former is not. This leads to four driving mode cases:

1. A train equipped with the old ATP/ATO equipment following another train equipped with the old ATP/ATO equipment will follow the existing rules and leave two un-occupied track circuits between them.
2. A train equipped with the *CITYFLO* 450 ATP/ATO equipment following a train equipped with the old ATP/ATO equipment can advance up to the end of the track circuit before the one occupied by the previous train.
3. A train equipped with the old ATP/ATO equipment following a train equipped with the *CITYFLO* 450 ATP/ATO equipment must also follow the existing rules and leave two un-occupied track circuit between them.



4. A train equipped with the *CITYFLO* 450 ATP/ATO equipment following another train with the *CITYFLO* 450 ATP/ATO can use the Moving Block capability in the CBTC.

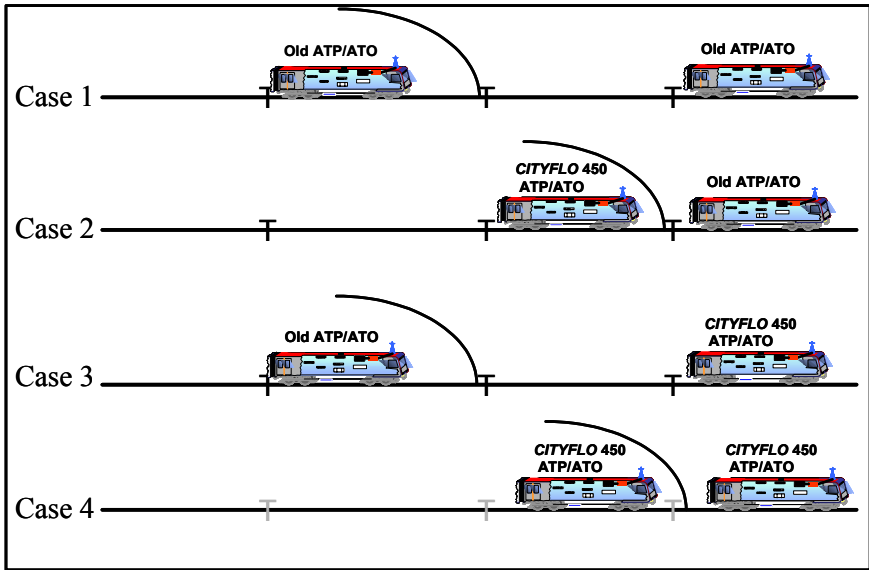


Figure 4: The four driving mode cases during system installation.

4.2.4 Meeting the headway requirement

If the dynamic headway requirement set up is 40 seconds, i.e. with a dwell-time of zero seconds in each station, the trains working in CBTC moving block mode would be separated by 40 seconds.

A modern CBTC system with moving block operation where the train 'footprint' or part of the track considered to be 'occupied' by the train is the train length plus a speed dependent 'buffer' area around the train. Moving and variable block, by its nature allows the trains to circulate closer to each other than fixed block system. The *CITYFLO* 450 system with its efficient radio communication has a demonstrated dynamic headway of about 15-20 seconds in other projects. Taking into consideration, the track and rolling stock characteristics of a typical metro it is nearly always possible to demonstrate in an operational simulation that the required dynamic headway will be met with *CITYFLO* 450 system.

4.2.5 Meeting the requirement for upgrading to a driverless system

The *CITYFLO* 450 CBTC moving block system with a driver is a version of the *CITYFLO* 650 CBTC moving block driverless system belonging to the same system family. It is therefore relatively easy to upgrade the signalling system itself to driverless operation.

Signalling systems have been capable of operating trains without driver for over two decades now. The issue is the compatibility of the infrastructure, i.e.



stations, tunnels and trains for driverless operation, especially the perceived security aspect and emergency procedures. Significant investment will be required for upgrading an existing metro system in order to adapt to the operational requirements of a driverless system.

5 Brown field Installations, a challenging implementation

For many European metros it will be their first experience in installing CBTC on an existing infrastructure although several CBTC systems have been installed in other metros in new lines. In such an installation we normally have to deal with a number of challenging issues such as:

1. The *CITYFLO 450* system uses Radiax cable as a medium for train to wayside communication. The cable can be mounted between the tracks, on the tunnel wall or overhead. While mounting the Radiax hanging from the tunnel roof between the two tracks would be logical, in many cases it is not considered feasible due to access restrictions, and instead the location is often changed to the tunnel wall. This requires two sets of Radiax cables to be laid – one on each sidewall impacting the cost and schedule of installation.
2. During the proposal phase, it is often envisaged to interface the onboard ATC equipment to two, three or more different types of trains. On detailed survey, it is often revealed that the interfaces are not uniform even on each type of train, which certainly increases the scope and complexity of the adaptation task.
3. The biggest challenge that awaits a project team is implementation of the system without affecting the existing operation. Most metros, operates 20 hrs each day and therefore only a short time window is available to access the track to perform installation activities. A meticulous planning, rigorous project management and strict organisational regime will be necessary in order to exploit this short access effectively.
4. Our experience on delivering similar application of CBTC to the LRT project in Philadelphia (SEPTA), USA suggests that the driver training for any new system is an onerous task. While, the normal train drivers are familiar with ATP/ATO operation, they will have to be trained in using the new operator console. The effort required in educating the different operational modes of this implementation, i.e. driving under CBTC (without signals), and under ‘fall-back’ mode (with signals), can not be underestimated.

6 Conclusion

The paper has demonstrated that a modern CBTC system with moving block capability and using modern computerized interlockings can be ‘overlaid’ in order to increase the transport capacity (throughput) of an existing Metro or Mass Transit line while maintaining the safety and the capacity of that line during the installation process.



Driverless CBTC – specific requirements for CBTC systems to overcome operation challenges

M. P. Georgescu

Alcatel Canada Inc, Transport Automation Solutions, Train Control, Canada

Abstract

Reliable driverless operation requires specific features implemented at system and subsystem levels of the train control system. Communications-Based Train Control (CBTC) is now proven as the best choice for driverless systems due to inherent high levels of safety and reliability with a low life cycle cost.

This paper proposes a systematic approach that may be used to determine the most efficient way to fulfil the requirements specific to each customer faced with driverless operation (green field or re-signaling). It also defines “must have” requirements (functionality) to obtain the desired performance and cost.

The paper also addresses issues related to the operability, maintainability, and availability of different types of driverless CBTC systems implementations, and the advantages and disadvantages of each solution.

Keywords: driverless, UTO, DTO, STO, CBTC.

1 Driverless operation with Automated Train Control

Industry members now unanimously recognize that the hierarchy for automated train control (ATC) systems is defined as Unmanned Train Operation (UTO), Driverless Train Operation (DTO), and Semi-automated Train Operation (STO) equivalent to the definitions in IEC 62290-1 (GAO2, GAO3, GAO4). For the purpose of this paper, both UTO and DTO will be expressed as Driverless operation. CBTC is defined as: Determination of train location, is to a high degree of precision, independent of Track Circuits; Continuous, bidirectional, train to wayside and wayside to train data communications, which can provide significantly more control and status information than is possible with a track



circuit based system; Wayside and train-borne vital processors to process the train status and control data and provide continuous automatic train protection (ATP), Automatic Train Operation (ATO) and Automatic Train Supervision (ATS) functions.

Driverless systems include sophisticated central control (i.e. ATS or System Management Centres) requiring few operators as many functions are automated.

STO has a driver in the cab who is responsible for safe departure of trains from stations, while the train drives automatically between stations, often with automatic turnback functionality. The driver observes the guideway for any hazardous condition, and will stop the train if a hazardous situation is observed. Train acceleration and braking is automated and the speed is supervised continuously by the system. The driver communicates with passengers and is part of the failure recovery strategy.

With DTO, a driver is not required in the front of the train to observe the guideway for hazards. There is, however, operations staff on board. Safe departure of the train from the station, which includes train door closing, can be the responsibility of the operations staff or may be done automatically.

With UTO, operation staff is not required on board the train. Safe departure of the train from the station, including door closing, is done automatically. As well, additional systems (such as Guideway Intrusion Detection, and platform and on-board CCTV) are usually installed to support the detection and management of hazardous conditions.

2 Driverless CBTC best fulfils train control operators' needs

Ideally, a train control system provides (proven) safe operation, low initial capital expenditure with the best value (quality/price) for each subsystem, increased system availability/reliability, and low operating and maintenance (O&M) costs, while providing increased operational flexibility (present and future extensions). The needs of the customer are best met by the Driverless CBTC solution.

2.1 Driverless implementation impact on subsystems

To implement driverless operation, design changes to several key traditional train control and auxiliary subsystems are required.

2.1.1 Signaling

This subsystem is crucial in providing system safety with high availability and reliability. It is also a major determinant to allow low O&M cost; the signalling subsystem provides flexible operation, support of failure management, energy savings, low maintenance cost, and low operation cost due to reduced staff.

2.1.2 Rolling stock

It is critical that fault indications, particularly with regards to safety, are communicated in real time to the central control operator, not to a driver. The



typical faults that are supervised by central (mandatory condition), are: dragging brake, motion obstructed, fire/smoke, low air pressure, park brake, slip-slide, doors status and train integrity.

Train Management System (TMS)/Health Monitoring Unit (HMU) information may be transmitted to central control and/or the maintenance shop via a separate non-vital communications link, not necessarily in real time, but still in a timely manner to trigger required unscheduled maintenance activities. It is acceptable that such information is sent via intermittent communication, e.g. downloaded from the train at each station.

“Driverless” implies automatic train control of propulsion and separate control for service, park and emergency brakes including revocable EB. Variable brake rates are implemented by the vehicle at ATC request.

Automated Speed Control is no longer a driver attribute. Sophisticated algorithms ensure jerk limitation and ride quality together with adhesion (slip-slide) monitoring and compensation.

Door control without train operator involvement, door recycling, as well as reaction to ‘door status lost,’ will also change. This requires on-board equipment to vitally supervise indication of propulsion disabling and brake status to allow the doors to open.

Redundant on-board ATC equipment should be installed. This includes remote reset capability used to restore redundancy, in a timely manner.

Driverless operation may reduce the fleet size, due to removal of the driver’s cabin, and shorter headway operation, which permit shorter trains (smaller fleet) to be used for optimum passenger service.

2.1.3 Wayside and trackside

2.1.3.1 Equipment. Traditional ancillary equipment is not required for true CBTC systems. This has a direct impact on capital cost, reliability/availability and O&M cost. With true CBTC systems, ‘alternate train presence detection’ subsystems, based on track circuits or axle counters become an operator choice, not a requirement.

2.1.3.2 Failure management. Fault tolerant technology ensures high availability, which is obtained with (two-out-of-three) 2oo3 configurations that still present a cost effective solution.

ATS failure is very rare, due to redundancy and highly available components provided by today’s technology. Without ATS the operator can retain fully automatic driverless operation

Wayside ‘auto-restart’ ensures automatic recovery following an unlikely failure of a wayside unit, to provide restoration of fully automatic operation within minutes. This sophisticated function has proven more efficient than any manual moves using “fall-back signaling” during years of reliable operation of SelTrac CBTC applied to AirTrain JFK (International Airport) and Detroit’s DPM.



2.1.4 Depot

Automatic operation in the depot should be considered mandatory to take full advantage of driverless operations.

Automatic routing of trains to and from storage areas, together with 'Storage Mode' for trains parked in a storage lane, and automatic 'wake-up' service to switch on necessary equipment, allows for seamless fleet build-up and reduction. This can happen any time in the operating cycle as a result of a scheduled transition, or can be triggered by Central Operator intervention as a response to an incidental change in passenger load demand. A by-product of driverless is the virtual extension of the depot in any spur and tail tracks that becomes possible, resulting in quicker changes in levels of operation.

Automatic close-up and shuffle-up in storage lanes and automatic coupling/uncoupling are powerful tools for the operator. This is important for efficiency of the operation with direct impact on energy savings and unscheduled rolling stock maintenance.

Car Wash for driverless trains travelling at low speed (e.g. 2 km/hr) can occur at anytime and be executed in record time.

2.1.5 Guideway

2.1.5.1 Infrastructure. If correlated with driverless CBTC, the civil guideway design can be optimized to achieve significant cost savings in the capital investment. Running shorter trains at lower headway has a positive impact on platform length determination.

CBTC allows for shorter end-of-guideway safety distances, which again lowers capital investment. New sections of guideway can be designed to allow significant cost savings.

Rear station turnback in conjunction with guideway and vehicle dependent safety distances results in shorter tunnels at the end of guideway.

Front station turnback design will further decrease infrastructure costs. In conjunction with FAO and an efficient alternate platform routing algorithm, the headway will remain the same as achieved by the rear turnback.

2.1.5.2 Guideway intrusion protection. Although we see them implemented in all types of train control systems, not having a driver in the cab mandates the installation of Guideway Intrusion Devices that interface with CBTC or Platform Doors subsystems.

2.1.6 Communication Subsystems (COMMS)

With traditional train control systems the driver provides the information and communicates with passenger in case of delays, changes of train travel, or emergencies.

For driverless operation, the communication system plays an important role and should provide the passengers with all required information, mainly in emergency situations. Due to the absence of a driver, the passengers' safety requires that the train and the wayside be equipped with a highly reliable communications system to provide on-board and platform voice information



(e.g., next station, delays, emergency instructions in case of fire, evacuation) and emergency intercoms that allow the passengers to communicate with the central operator, on top of typical platform displays of arrival time and destination.

A CCTV subsystem becomes mandatory on the platforms and optional inside the trains.

Today’s technology of voice-over IP and increased bandwidth allows for full integration of COMMS and/or CCTV functions in the ATS network.

2.2 Operations and maintenance impact

2.2.1 Operating costs

Although new jobs are created for passenger service and maintenance, CBTC facilitates lower operational costs because the drivers and their associated ‘wear and tear’ on the vehicle are eliminated. On top of the overall optimal system performance, the cost of driver employment, while running shorter trains, can be significant – up to about 14% of the operations cost as presented by Prof. H. Strobel [2].

The overall operating costs are lower. Vancouver SkyTrain continues to save millions with the most efficient sized work force in the industry [1]. Direct Operating costs per passenger kilometre released by APTA 1989 [1] are presented in Figure 1.

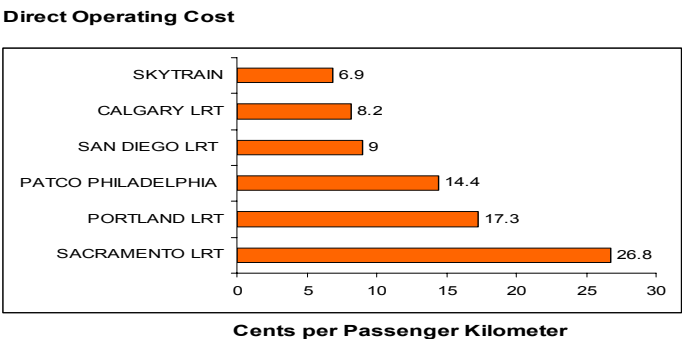


Figure 1: Operating costs per passenger kilometre.

With the cost energy increase while extending the number of stations from 16 to 32, the SkyTrain Operating cost per passenger kilometre in 2005 was 10.15 cents, way lower than Philadelphia 1986 number,

2.2.1.1 Facilitates energy optimization. With driverless operation, the optimization of regenerative braking presents the possibility of the kinetic energy of the train to be recovered during braking, by allowing train braking into a station to be coordinated with train acceleration out of another station. The regenerative energy can be consumed by the accelerating train. As stated in BC Transit’s Fact Sheet – SkyTrain Performance [1], Vancouver SkyTrain “Energy consumption per passenger-kilometre is just less than half the average of the

other systems” Figure 2 illustrates the energy consumption for North American transit industry [1] (using APTA 1989 data).

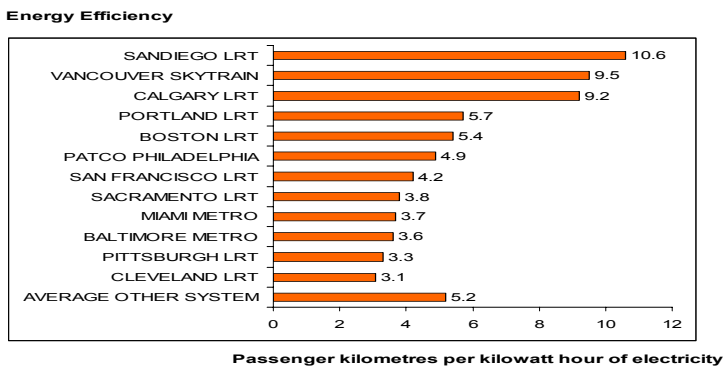


Figure 2: Passenger kilometre per kilowatt hour of electricity.

In 2005, SkyTrain “Passenger Kilometre per Substation Kilowatt Hour was 6.8. This number is used to establish the traction power costs and does not include the yard, maintenance facility and stations.

2.2.1.2 Increased operational flexibility. Automated operation allows central operating staff to concentrate on any emerging problems. By design, while train performance is continuously controlled and monitored, driverless CBTC systems automatically maintain an efficient schedule. Liberated from these tasks, central control operators can concentrate on fast response to unusual situations not covered automatically by the system. Better operators’ decision support is provided by intelligent CBTC systems, which can predict future conflicts long before they occur. Therefore conflicts can be avoided using the tools provided. Optimum recovery from major failures is implemented quickly. A perfect example is managing gaps in service, e.g. running supplementary unscheduled trains in parallel with scheduled trains or replacing faulty trains in service

2.2.1.3 Capacity. Firstly, capacity is increased by more frequent service of shorter trains instead of less frequent service by longer trains. Secondly, attractive options during non-peak hours are available by running variable length trains while maintaining an acceptable level of service.

3 Driverless operation past, present, future

The demand for driverless CBTC trend since 1981 is presented in Figure 3. The number of CBTC applications soars between 2001 and 2006 and will remain on the ascending curve.



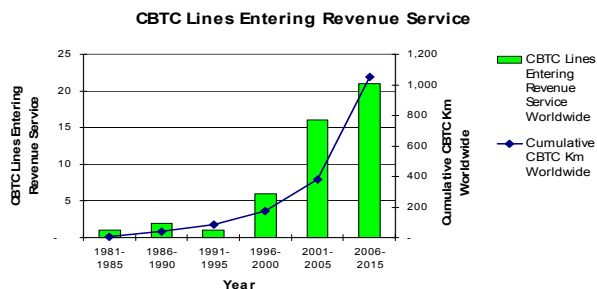


Figure 3: CBTC lines entering revenue service.

The split between re-signaling and green field projects presented in Figure 4 appears to indicate the transit operators' preference for the type of control (illustrated in Figure 5) at the beginning of 21st Century. Basically most of the operators going through re-signalling with CBTC appear to prefer to keep a driver in the existing train cab, at least for the initial life of the renewed system. Green field applications options prefer UTO fully automatic operations to take advantage of the best value transit systems available as described in Moving to Full Automatic Operations [3].

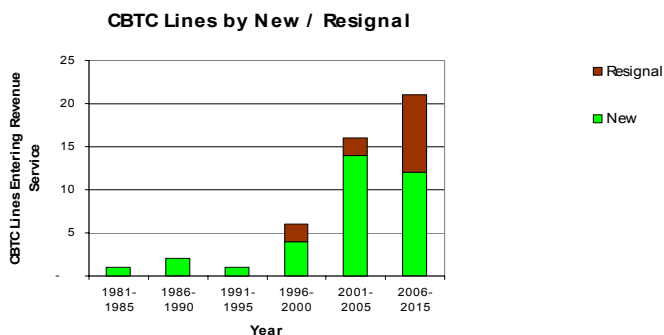


Figure 4: CBTC Lines by new / resignal.

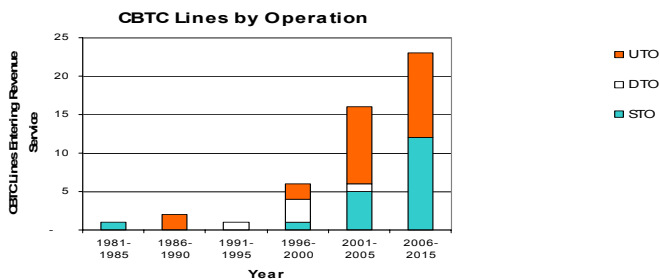


Figure 5: CBTC lines by operation.



Since 1981, 46 CBTC projects have been awarded with a total of 940 Km, as illustrated in Figure 6. From this total, 58.69% are driverless. 33.33% of re-signaling projects selected DTO or UTO while 29% of green field projects are implementing STO.

Customer & Line	Signaling Company	Type of Service	Operation	New / Resignal	Award Year	Length Km
Toronto Scarborough	Alcatel	Metro	STO	New	1981	6.4
Vancouver Sky Train	Alcatel	Metro	UTO	New	1981	49.7
Detroit APM	Alcatel	APM	UTO	New	1983	4.7
Lyon	Siemens	Metro	DTO	New	1986	45.0
London DLR	Alcatel	Metro	DTO	Resignal	1992	31.0
Paris RATP Line 14 Meteor	Siemens	Metro	UTO	New	1992	7.0
San Francisco MUNI	Alcatel	LRT	DTO	Resignal	1992	11.0
Ankara	Alcatel	Metro	STO	New	1993	14.6
Kuala Lumpur	Alcatel	Metro	UTO	New	1995	29.0
Singapore North East Line	Alstom	Metro	UTO	New	1997	20.0
Hong Kong KCRC West Rail	Alcatel	Metro	STO	New	1998	30.5
New York City JFK	Alcatel	Metro	UTO	New	1998	13.0
San Francisco BART	GETS	Metro	STO	New	1998	12.0
San Francisco Airport	Bombardier	APM	UTO	New	1999	9.6
Seattle Airport	Bombardier	APM	UTO	New	1999	2.7
New York City Transit Canarsie	Siemens	Metro	STO	Resignal	1999	17.0
Las Vegas Monorail	Alcatel	APM	UTO	New	2000	6.4
Dallas/Fort Worth Airport	Bombardier	APM	UTO	New	2000	7.7
Philadelphia SEPTA	Bombardier	LRT	STO	Resignal	2000	4.0
Hong Kong KCRC Ma On-Shan	Alcatel	Metro	STO	New	2001	11.4
Nurnberg - Metro resignal	Siemens	Metro	UTO	Resignal	2001	21.5
Singapore Marina Line	Alstom	Metro	UTO	New	2001	10.6
Wuhan LRT	Alcatel	Metro	STO	New	2002	10.2
Hong Kong MTRC Disney Resort Line	Alcatel	APM	UTO	New	2002	3.2
Paris RATP Line 13	Alcatel	Metro	STO	Resignal	2002	26.0
Lausanne Metro	Alstom	Metro	UTO	New	2002	6.0
Singapore Circle line Stage 3	Alstom	Metro	UTO	New	2002	19.0
Barcelona Line 9	Siemens	Metro	UTO	New	2003	42.0
Seoul KNR Bundang	Alcatel	Metro	STO	Resignal	2003	18.3
Guangzhou Metro Line 3	Alcatel	Metro	STO	New	2003	36.0
Taipei Neihu Line	Bombardier	Metro	UTO	Resignal	2003	15.0
London LUL Jubilee	Alcatel	Metro	STO	Resignal	2003	35.0
London LUL Northern	Alcatel	Metro	STO	Resignal	2003	67.0
Guangzhou L4L5 - Metro new line	Siemens	Metro	STO	New	2004	69.0
Madrid Line 1&6- Resignalling	Bombardier	Metro	STO	Resignal	2004	40.2
Paris RATP L3 - RESIG	CSEE	Metro	STO	Resignal	2004	12.0
Paris RATP L5 - RESIG	Siemens	Metro	STO	Resignal	2004	14.5
London Heathrow Airport People Mover	Bombardier	APM	UTO	New	2004	1.0
Washington Dulles Airport PM (Ph I)	Alcatel	APM	UTO	New	2004	3.6
Yong In	Bombardier	APM	UTO	New	2004	18.0
Shanghai Yangpu Line (M8)	Alcatel	Metro	STO	New	2005	23.0
Budapest	Siemens	Metro	STO	Resignal	2005	10.5
Dubai Light Metro - Red Line	Alcatel	Metro	UTO	New	2005	49.0
Paris Line 1	Siemens	Metro	UTO	Resignal	2005	16.5
Beijing Airport People Mover	Bombardier	Metro	UTO	New	2005	1.9
Vancouver Canada Line	Alcatel	Metro	UTO	New	2006	18.0

Figure 6: CBTC Projects Awards 1981 – 2006.

4 Conclusion

With more than 20 Urban Rail Driverless Systems in revenue service worldwide and the increasing demand for driverless operation for green field and re-signaling projects, the STO, DTO, UTO is an operator choice not a technology constraint.

Regardless of the solution selected, a systematic approach is required to assess customer needs. In particular, driverless operation has specific mandatory



requirements reflected in “must have” functions to ensure reliable and efficient operation. Driverless CBTC meets today’s operator challenges of increasing performance and passenger service.

References

- [1] “BC Transit Fact Sheet – SkyTrain Performance”, October 1991
- [2] H. Strobel, J. Schuette , Presentation to UITP Conference on Automated Metros, Nurnberg 2004
- [3] Firth Whitwam, Mircea Georgescu - Moving to Full Automatic Operations, IEEE Hong Kong 2005



This page intentionally left blank

Section 7

Advanced train control

This page intentionally left blank

CBTC (Communication Based Train Control): system and development

N. Bin, T. Tao, Q. K. Min & G. C. Hai

*Department of Control Engineering,
School of Electronics and Information Engineering,
Beijing Jiaotong University, Beijing, People's Republic of China*

Abstract

CBTC (Communication Based Train Control) systems are known as comprehensive, integrated and intelligent control systems for rail systems including mainline railways, light rails and underground lines in cities. With the development of modern data communication, computer and control techniques, CBTC represents the future direction of rail control systems. At present, CBTC has been used in light rail and underground lines in cities. It has not been implemented in mainline railways for many reasons. In future decades, rail systems will be in rapid development periods throughout the world. CBTC is known as the brain and nerve centre of rail systems, and ensures the safety and efficiency of rail systems. It is necessary for CBTC to be researched and developed further. In Europe, there is the ETCS (European Train Control System). In China, there is the CTCS (Chinese Train Control System). In Northern American and Japan, there are advanced train control systems or moving block systems. However, there is no standard for development and design of CBTC in the world at present. In this paper, efforts are made towards the establishment of a CBTC standard which directs the development and design of CBTC systems. The configuration of CBTC systems is first described. The key technical issues are addressed. The fundamental modular of CBTC and its interface requirements are defined. The transit methods from the present train control systems based on track circuits, transponders and other traditional means to CBTC systems are also put forward.

Keywords: automation control system, rail system, computer and communication, standardization.



1 Introduction

The CBTC (Communication Based Train Control) system has been known as the development direction of control systems for rail systems in the world. In particular, with quick development of modern mobile communication, its implementation and application become more and more easily. At present, CBTC has been used in city rail transportation systems, such as light rail and underground systems. It will be used in mainline railway systems in the near future. Application of CBTC has the following features. It makes the dispatching system more flexible and efficient. The safety and reliability of the system are high. It is easy for CBTC to be transited from the present system. The maintenance cost for CBTC is lower since track circuits are removed from the system. The control system of railway network will be towards intelligent, network and comprehensive system, CBTC is as brain and nerve centre of the railway system. Its development and application will be with the direction of railway operation control system [2].

At present, there are CBTC systems from the different company in the world. For example, SELTRAC from Alcatel has been used in SkyTrain in Vancouver in Canada since 1986, in JFK-Airport Light Rail System in the United States since 2003. Trainguard MT from Siemens will be used in the underground line Canarsie in New York in 2006. URBALIS from Alstom has been used in the Light rail in Singapore since 2003. In addition, there are also CBTC systems from Japan Signaling, GE and CSEE etc. In the world, there are more than 30 light rail and underground systems where CBTC systems have been or will be applied. According to the statistics, most of the CBTC systems are based on cross-loops for train-ground communication, some of the CBTC systems use radio for train-ground communication. In Europe, Radio system will be used for train control (ETCS-2 and Euro-radio) [1]. Nowadays, when the signaling systems need to be upgraded in many cities, such as London, Paris and New York, CBTC system is chose as a new system. It is predicted that CBTC systems will be applied in mainline railway in the near future [5].

However, there is no standardization for CBTC systems. The CBTC systems from the different company cannot be compatible. It is easy for a kind of CBTC system from a company to monopoly the market. It is not good for commercial competition and technical development. For users of a railway network, it is not possible to select the best systems. Meanwhile, it is not easy for the CBTC system to be upgraded with technical development. In the paper, efforts are made to put forward the technical standardization for CBTC in terms of system configuration, function requirements, data format, interface definition and development in order to facilitate development and design of CBTC systems. Meanwhile, CBTC systems from the different companies can be compatible on a railway network and railway users have more choices in the CBTC markets. Of course, for every designer of a CBTC, it is easier for the whole system or part of the system of a CBTC to be upgraded with new technology advent.



2 System configuration of CBTC

A CBTC system can be divided into the five parts. The first part is the Central Control System (CCS). The second part is the Station Control System (SCS). The third part is the Onboard Control System (OCS). The fourth part is the Block Control System (BCS), including Radio Block Control System (RBCS) and block sensors etc. The fifth part is the Communication Network System (CNS), including mobile communication system. The figure 1 shows the configuration of a CBTC system. The five parts constitutes the whole system of a CBTC system. However, the five parts are relatively an independent part each other as a subsystem of a CBTC. The fifth part (CNS) connects all other four parts as communication channel, including mobile communication between train and wayside systems in order to ensure real-time, reliable and safe data exchange among them. The functions of a CBTC system should be distributed to the five subsystems. The data format between every two parts should be defined.

The Central Control System (CCS) is the control center of a CBTC system. Train plans and train graphs are generated here. All the train operation is dispatched and commanded in CCS according to train graphs. The state data concerning to station control system (SCS), block control system (BCS), on-board control system (OCS) and train operation should be sent to CCS [1].

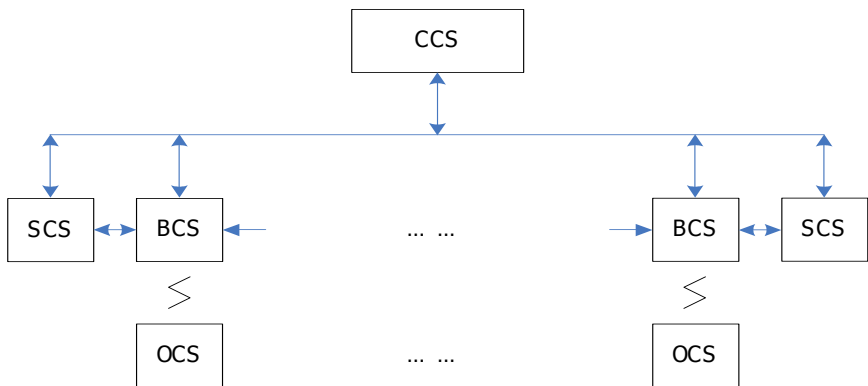


Figure 1: The configuration of CBTC systems.

Station Control System (SCS) is an interlocking system which control switches, signals and routes at stations or in areas. SCS communicate with CCS, BCS and OCS. Station interlocking system is a traditional train operation control system in terms of its functions. There is no special requirement for SCS in CBTC systems. At present, most of SCS are the computer based interlocking systems in CBTC systems [2].

Block Control System (BSC) includes radio block system and wayside sensors for CBTC system along the track. BSC communicate with CCS, SCS and OCS. Its main function is to control train safety operation in blocks. Train operation safety interval is calculated by BCS according to train safety operation



modes on moving block conditions, and then the concerning orders (permissive speed) are sent to OCS by CNS.

On-board Control System (OCS) is a train control system equipped in locomotives. Its main function is to control train speed such as acceleration, deceleration, cruising and braking. Data concerning train position and speed are sent to BCS. The permissive speed of train operation is received by OCS [3]. In a CBTC system, OCS must be safer and more intelligent compared with OCS in the traditional train control system.

SCS, BCS and OCS must be failsafe and reliable. CNS must be satisfied with the requirements of real time, safety and reliable data transmission in CBTC systems. For CCS, there is no failsafe requirement since it does not directly control train operation, but it is reliable [3].

As a comprehensive train operation control system, it is reasonable for a CBTC system to be divided into the five subsystems. Every subsystem is relatively independent and easier to be designed and implemented. The above division is very helpful for a CBTC system to be analyzed, designed and developed.

3 The key technical issues and interface requirements of CBTC

In a CBTC system, the key technical issues can be described as follows. Some of the key technologies are common in every subsystem. The others are relatively independent.

The vital computer is one of the core parts in CBTC systems. In SCS, BCS and OCS, there is a vital computer. From the point of view of application, the vital computer is different with the different subsystems. However, in terms of safety and reliable requirements, the safety platform of the vital computer in the different subsystems is the same. The vital computer can be designed as a series modes which are suitable for different subsystem in a CBTC system. The configuration of a vital computer can be two out of two or two out of three where fault-tolerant design is carried out. The safety platform is transparent to the application in the different subsystems. The vital computer is relatively independent parts in CBTC system. Its configuration and software platform can be upgraded with technical development of computer and fault-tolerant design [6].

Figure 2 shows the configuration of an OCS in a CBTC system. The vital computer is the core part of OCS. Other parts (I, II, III, IV...) are connected to the vital computer by a kind of bus (Can bus or other filed bus). Other parts could be radio receiver and transmitter unit, position unit, speed measurement units, MMI unit, locomotive engine interface, recorder unit etc [3]. It is obvious that OCS is modular configuration, in addition to the vital computer.

The reliable and safety mobile communication system is the foundation of a CBTC system. The concept about moving block system (CBTC is also called as moving block system) was put forward in 1960. It is not possible for a CBTC system to be implemented until the reliable, safe mobile communication appears.



Traditional track circuits cannot satisfy with the communication requirements of train and wayside in a CBTC. In the last decades, cross-cable and leakage cable or leakage optical fiber can be used in CBTC systems. Today, GSM-R and other radio system begin to be applied in a CBTC system. The reliable and safety mobile communication system become the key technology of a CBTC system.

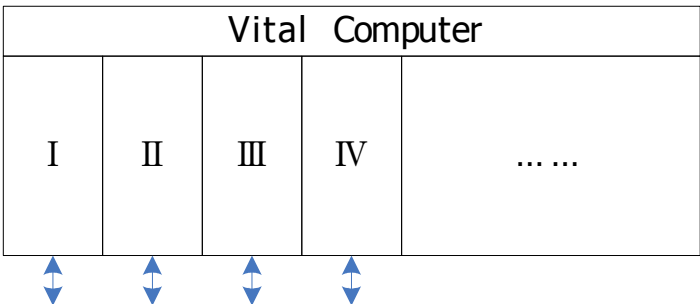


Figure 2: On-board control system configuration.

The accurate position system and the accurate speed measurement system are also the key technologies of CBTC systems. The accurate position and speed is the basic parameters of train operation control. In the position system, position calibration must be considered. In the speed measurement system, it is possible for combination of the different speed sensors (radar and axle generator) to be used to be suitable for the different speed of the train [3].

Train integrity system is very important unit in a CBTC system since track circuits are removed. An axle-counting system or on-board train integrity unit are applied for train integrity checking in a CBTC system.

The dispatching algorithms are the core software in CCS of CBTC systems. Its task is to generate train graphs according to the requirements of train plan and to automatically restore normal train operation when train operation plan is disturbed.

The train operation control model is the key to ensure train to be safe operation in CBTC systems. In a CBTC system, there is no block section. Train following interval is calculated in real time method. The figure 3 shows the principle of train following interval control in CBTC systems. In addition to the safety protection distance (d_f), the interval of the two following trains (train 1 and train 2) is the safe braking distance of train 2 in theory. $V_2(d)$ is the speed curve of the following train 2. $V_{2e}(d)$ is the emergency braking curve of the train 2, and $V_{2s}(d)$ is the service braking curve of the train 2. O is the calculated stop point of train 2. O' and O'' are respectively the actual stop point of train 2, caused by the various errors. It is obvious that train operation control mode is the algorithm of train speed and interval control. Cellular Automata model is the newest model for train operation control [4].

Simulation and test platform is necessary for CBTC system development. Since CBTC system is a complicated and comprehensive system, during the development of a CBTC system, it needs a simulation environment to support

the development. After the CBTC system is implemented and put into operation, it needs a simulation platform to test and maintain it.

4 Development, implementation and transit

CBTC is a comprehensive system over a rail network. The following principles should be observed in the development and design of a CBTC system.

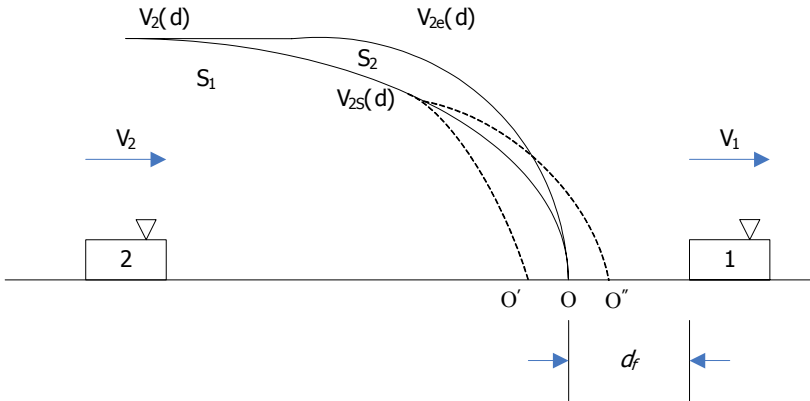


Figure 3: Train Interval control model.

According to the requirements of mainline railway network in an area or a country, light rail network or underground network in a city, the system requirements specification (SRS) and system functions specification (SFS) must be made up. These are the first set of files for a CBTC system, and they are also the basic files for a CBTC system. The files could be different for new lines and old lines upgrading in a network. This is the first step of the CBTC system development.

The second step is the key technical selection of a CBTC system. It includes the communication technology between train and wayside, the type of the vital computer and the system configuration. For communication system between train and wayside, it could be GSM-R, 2.4GHz Direct Sequence Spread Spectrum (DSSS or Frequent Hoping Spread Spectrum-FHSS) system in IEEE802.11 or cross-cable system etc. For the type of vital computer, there are more choices. It could be two out of two system or two out of three systems based on the different commercial computers. Decision on the above technologies is closely related to the cost and reliability of the CBTC systems.

Interface standardization and data format standardization should be defined before the system design. Firstly, the interfaces between CCS and SCS, CCS and BCS, SCS and BCS, OCS and BCS are defined. The data format transmitted among the subsystems and their contents are defined. Development of each subsystem will be relatively independent and the whole system will be modular. For users, the configuration of the CBTC system will be flexible.



After the above design files are finalized, system design and development can be started. Files management and design steps must be line with the requirements of software engineering and safety assessment procedures since a CBTC system is a safety and reliable system. Design and development for subsystems can be carried out in parallel. The chief designer must coordinate the progress of each subsystem [6].

In order to promote and verify the design and development, simulation test will be carried out in the whole process of the design. Normally, a general simulation environment platform is established to test each subsystem. Finally, after the whole system is finished, it should be simulated in laboratory to test its functions and safety features in varied conditions before it is installed in the field.

Before a CBTC system is put in operation, field test must be carried out since the simulation test in laboratory cannot be perfect. Particularly, the physic conditions such as temperature, humidity, vibration and electro-magnetic interference etc are different. The period of field test depends on the laboratory simulation contents and the field conditions.

According to the development of rail transportation system and CBTC systems in the world, it is only the problem of time for the traditional train control system (track circuits based train control system) to be replaced by CBTC systems. At present, CBTC systems have been applied in light rail network and underground systems in the world. Particularly, when the signaling system is upgraded and the new line is constructed, CBTC is their first choice for train control systems. For example, the New York Metro selected Siemens's CBTC and Shanghai Metro selected Alstom's CBTC systems. It is also clear that CBTC will be applied in main line railway network. In Europe, ETCS level 2 has been in trial since 2000 [5].

It is true that transit from traditional train control system to CBTC systems is a long process. It is not possible for the transit to be completed in the short time since the cost and other reasons. However, from the above CBTC system analysis, it can be seen that the transit from the traditional system to CBTC system is very easy. This is one of the biggest advantages of CBTC systems since that can overlap on the traditional systems. Under the certain conditions, the two systems can coexist without interference. Before the CBTC is adjusted into normal operation, the old system can still be used. The traditional system can be removed only after the CBTC is in normal operation.

Since the CBTC system is modular system, the part of the traditional system can remain in the CBTC system. For example, as a relatively independent subsystem, the interlocking system can still be used in the CBTC system as long as the interface is upgraded. Therefore, when signaling system needs to be upgraded, CBTC can be implemented step by step.

5 Conclusion

It is very clear that CBTC system will replace the traditional train control system in railway network in the world. Its configuration, interfaces and data format



should be standardized. Its key technical issues are in common. As long as the standardizations can be observed, the CBTC systems manufactured by the different companies can be compatible on railway network. Moreover, CBTC system can be easily upgraded with development of new technologies in the key technical areas. Meanwhile, it is also easier for the traditional train control system to be transited into CBTC systems.

References

- [1] Bin Ning “ETCS and CTCS “ Computers in Railways IX WIT Press pp 262 – 272 May 2004, Comrail 2004, Drensdon, Germany.
- [2] Bin Ning “A comprehensive Information System for Railway Network” Computers in Railways VIII, WIT Press, pp 152 – 162, June 2002, Comrail 2002, Lemnos, Greece.
- [3] Bin Ning “Intelligent Railway System in the 21st Century” Computers in Railways VII, WIT Press, pp 1153 – 1163, September, 2000, Comrail 2000, Bologna, Italy.
- [4] Bin Ning, K.-P. Li and Z.-Y. Gao “Modeling Fixed-Block Railway Signaling System Using Cellular Automata Model” International Journal Of Modern Physics C, Volume 16, Number 11 November 2005 pp 1793 – 1801.
- [5] Eddie Goddard “London Underground Signaling Renewal” IRSE News Issue 111, February 2006. pp 1 – 7.
- [6] Rail Safety and Standards Board “Engineering Safety Management” (the “Yellow Book”).



An algorithm for braking curve calculations in ERTMS train protection systems

B. Friman

*Friman Datakonsult AB and Uppsala University,
Human-Computer Interaction, Department of Information Technology,
Sweden*

Abstract

The European Railroad Transportation Management Systems standard for train protection, ETCS, includes several advanced features for predicting the safe speed from a number of target locations ahead of the train. The braking system can have a different braking capability in different speed segments. The area in front of the train can contain a number of targets with different target and release speeds. The area in front of it is also segmented according to the gradients, in a way which is independent of the targets. The variables that shall be input to the braking curve algorithm therefore have three dimensions. Since permitted speed shall be calculated, rather than time to intervention, square roots are needed for the calculations, which require some computational power. The article suggests an algorithm where the gradients and the targets are combined in one table. This makes the input variable area two-dimensional instead of three-dimensional, which simplifies calculations and reduces the necessary number of square root calculations.

Keywords: ETCS, ERTMS, braking curve, ATP.

1 Introduction

The European Train Control System – the new European standard for automatic train protection, uses basically a three-dimensional set of data as input to the supervision algorithms. The three dimensions are: a table of speed restrictions, a table of gradient sections, and a table of deceleration abilities in relation to speed. The output from the algorithms is expressed in the form of speed values; e.g. which is the highest possible speed the train can run, in order to be able to



obey all speed restrictions on the track ahead of the train? The three dimensions of data input may require a large number of calculations. The requirement for output to be expressed in the form of a speed, requires square roots to be used, which requires comparatively large computational power for each calculation, especially since real-time CPU'S are not normally equipped with dedicated floating point processors.

Enough computation power must be allocated to be able to supervise the maximum possible number of restrictions, gradients and deceleration segments, plus an ample safety margin, in order to satisfy the safety requirements of a train protection system. All calculations shall be repeated according to changes in input data, such as train position, train speed and when information about new speed restrictions becomes available.

This paper describes an algorithm using a two-dimensional approach for the calculation of the output speed values.

Section 2 will describe the basic requirements for train speed supervision in ETCS. In section 3, the two-dimensional approach and algorithm will be described. In section 4, additions to this algorithm to be able to handle various special ETCS requirements will be described. Finally in section 5, the conclusions of this work will be summarized.

2 Basic requirements for train speed supervision in ETCS

The following is a simplified description of the requirements for train speed supervision in ETCS. Beside the following requirements, there are also other requirements such as separate service and emergency brake supervision, consideration to position measurement uncertainty, and numerous other requirements which are needed in ETCS, but which are not necessary for describing the basic algorithm for supervision of targets ahead of the train. Possible algorithmic solutions to some of the more detailed requirements are however discussed later in this article.

2.1 Speed restriction table

The speed restriction table can contain up to 31 restrictions ahead of the train (30 static speed restrictions and one distant signal speed restriction). (See `n_iter` in packet 12 and packet 27, in subset-026 chapter 7 and in subset-58.) The restriction speeds can vary between 0 km/h and 600 km/h in steps of 5 km/h. (See variable `v_static` and `v_loa` in packets 12 and 27 in subset 026 chapter 7 and in subset-058).

2.2 Gradient section table

The gradient section table can contain up to 31 sections defining the gradients on the track in front of the train. The gradient on one section can be between -25.4 and $+25.4\%$ in steps of 0.1% (see variable `g_a` om packet 21 in subset-026 chapter 7 and subset-058).



2.3 Deceleration ability table

The deceleration table describes the trains' ability to brake, as a function of speed. The use of such a table makes it possible to brake later in certain situations, since the known differences in braking ability related to speed are taken advantage of. The table cannot be longer than 31 speed segments. Each speed segment defines the trains braking ability in a certain speed range. The braking ability can be between 0 and 2.55 m/s/s in steps of 0.01 m/s/s.

2.4 Brake delay time

Two models for brake delay are allowed. One assumes that the braking ability is zero during the brake delay and 100% after the brake delay. The other model uses two delay intervals. The braking ability is assumed to be zero during the first interval. During the second interval, braking ability is assumed to be gradually increasing to 100%. The algorithm described in this paper uses the first, simpler brake delay model.

3 A two dimensional approach and supervision algorithm

The supervision algorithm shall calculate a palette of speeds for various purposes, from the input data in form of speed restrictions, gradient sections, deceleration table and brake delay time. This palette includes:

- Service brake intervention speed SBI: Which is the highest speed that the train can run and still be able to obey all speed restrictions ahead of the train, using only the service brake?
- Emergency brake intervention speed EBI: Which is the highest speed that the train can run and still be able to obey all speed restrictions ahead of the train, using the emergency brake (which has significantly shorter brake delay time)?
- Warning speed W: Which is the highest speed that the train can run and the driver has still a few seconds margin before the SBI speed is exceeded (considering that the SBI speed becomes lower and lower when the train approaches a the start of a speed restriction)?
- Permitted speed P: Which is the highest speed that the train can run and the driver has still a few seconds margin before the W speed is exceeded (considering that the W speed becomes lower and lower when the train approaches a the start of a speed restriction)?

This article will describe an algorithm that can be used to calculate either of the above measures, depending on which brake delay time is used, which position uncertainty is added and which restriction margins are selected. A plausible strategy would be to first calculate the EBI and then use the same algorithm again to calculate SBI, W and P, this time focusing on the restriction which showed to be most restrictive for EBI. The calculation of SBI, W and P can be done in one pass, so a total of two passed would then be necessary, where only on restriction would be considered in the second pass.



3.1 Basic steps in algorithm

When calculating the trains braking ability towards a restriction, only the allowed speed and starting point of the restriction are of interest. These are in the following called targets. A target is thus a combination of a position and an allowed speed (beginning at that position). That implies that the restriction table is seen as a table of speed targets which the train must be able to brake to.

The first obvious step in the algorithm is to remove all targets which cannot possibly be the most restrictive target. That is the case when a more distant target allows the same or a higher speed than a closer target. After this first step, the target/restriction table will contain a down slope stair where each more distant target has a lower allowed speed than the previous. The first step is obvious and will not be further discussed in this article.

The second step in the algorithm is to merge the target/restriction table and the gradient table into one table which contains both target speeds and positions in the track where the gradient changes.

The third step in the algorithm is to calculate the brake delay distance, that is how long the train would run without any braking, given the train speed and the braking delay. Later in the article I will argue that why it is not the trains actual speed that shall be used to calculate the brake delay distance, but rather the previous result from using the algorithm (the highest possible speed which would make it possible to obey all targets in front of the train).

The fourth step in the algorithm is to calculate the allowed speed at all positions in the target/gradient table, starting with the last position and working stepwise backward towards the position of the train. At each step, a target speed to current allowed speed calculation is done. In this calculation, the gradient (which affects the trains braking ability) is fixed, since all gradient change positions are included in the table. The trains braking ability relative to speed may however be non-fixed. The trains' deceleration ability at the result speed may differ from the ability at the target speed. Therefore, this calculation is in itself divided into steps, one for each involved segment in the deceleration ability table. The resulting allowed speed shall be compare with the target speed at the new position if there is one and the lowest of both selected as allowed speed at that position. If there is not target speed at the new position (the new position represents a gradient change), then the result of the calculation shall be used as allowed speed at the new position.

The fifth step is to interrupt the calculations when the position becomes closer to the train than the brake delay distance, since the trains braking ability is assumed to be zero here. Targets on the distance where braking ability is zero, shall instead be directly compared with the calculated allowed speed, and the lowest value selected.

3.2 Step 2: merging the targets/restriction and gradient tables

The merging of the target/restriction and gradient tables can be illustrated by the following example:



Table 1: Example of a target/restriction table.

Position	Allowed speed
500 m	100 km/h
1000 m	50 km/h
1500 m	0 km/h

Table 2: Example of a gradient table.

Position	Gradient up to this position
700 m	0%
1200 m	-1.0%
1700 m	-2.0%

Table 3: Example of a merged target/restriction + gradient table.

Position	Type	Allowed speed	Gradient up to this position
500 m	Restriction	100 km/h	-
700 m	Gradient change	-	0%
1000 m	Restriction	50 km/h	-
1200 m	Gradient change	-	-1.0%
1500 m	Restriction	0 km/h	-
1700 m	Gradient change	-	-2.0%

3.3 Step 3: calculating the brake delay distance

The braking ability is assumed to be 100% closer the targets. But before a certain position, the braking ability is assumed to be zero. This position is calculated as the current position plus the brake delay multiplied by the train speed:

$$pos_{delayend} = pos_{train} + brakedelay \cdot v \quad (1)$$

For the sake of knowing when to issue an automatic ATP brake, the actual train speed can be used in this algorithm. If however the speed result from the calculation shall also be used as information to the driver, then we must use the resulting speed from the calculation when calculating the brake delay distance. A larger train speed will result in a larger brake delay distance which in its turn will result in a lower allowed speed from the algorithm. This means that there is a circular dependence in the algorithm. Is this a problem? Assume that the allowed speed is calculated periodically as the train advances on the track. When the train approaches a target, the allowed speed will gradually decrease. If the algorithm used the result from the previous calculation when calculating the brake delay distance, then the dependence on previous cycles result may in this case cause a minor underestimation of the allowed speed which is showed to the driver. The error can be approximated as the trains braking ability during the time between two calculations. If, for example, the braking ability is 0.8 m/s/s, and the calculations are done with 0.25 s intervals, then the error would be 0.7 km/h, and it would be in the safe direction. If for example a future restriction suddenly



disappears, for example because a signal ahead of the train changes from stop to clear, then however the calculated allowed speed would be too high, because it would be based on a too short brake delay distance, and the error could be rather large. This problem can however be solved by limiting how much the allowed speed may increase from one calculation to the next, or by delaying the display of the new information until one the calculation has been performed two times (this could for example result in a delay of the increased speed display by 0.25 s).

3.4 Step 4: calculating the allowed speed at all positions in the table

The basic formula for calculating the braking distance from one speed v_1 to a lower speed v_2 is:

$$dbr = \frac{v_1^2}{2 \cdot dec} - \frac{v_2^2}{2 \cdot dec}, \quad (2)$$

where dec is the trains' deceleration ability, in m/s^2 .

When dbr, dec and v_2 are available, v_1 can be calculated as:

$$v_1 = \sqrt{2 \cdot dbr \cdot dec + v_2^2} \quad (3)$$

So, if the allowed speed at position p is known, then the allowed speed at position p-1 can be calculated as:

$$v_{p-1} = \sqrt{2 \cdot (pos_p - pos_{p-1}) \cdot dec + v_p^2} \quad (4)$$

The trains' deceleration is dependent of the gradient and of the train speed. However, we know that the gradient over the distance is fixed and available as the next gradient change in the table (in tablerow (p) or later).

The relation between gradient and deceleration is defined by the following formula:

$$dec = dec_0 + g \cdot \frac{grad}{100} \quad (5)$$

where dec_0 is the zero gradient deceleration, $g \approx 9.8186 \text{ m/s}^2$ (free fall acceleration) and gradient grad is expressed in %. The formula can be easily understood if we consider the case when the dec_0 is zero and the slope is 100%. The deceleration would then become -9.8186 m/s^2 , which represents free fall acceleration.

As mentioned earlier, dec_0 is dependent on the trains' speed, in accordance with the deceleration ability table. If we assume that there is a function $f_{vto\text{dec}}$ available which looks up the table and returns the dec_0 deceleration as function of train speed, then we know that the dec_0 is equal to $f_{vto\text{dec}}(vp)$ at position p. dec_0 at position p-1 is however unknown at the moment, but we assume tentatively that it is the same as at position p, and can then calculate dec as

$$dec = f_{vto\text{dec}}(v_2) + g \cdot \frac{grad_p}{100} \quad (6)$$



Using dec above, v_1 is now tentatively calculated. If v_1 has the same deceleration dec_0 as v_2 , then v_1 is correct. To check this, we assume there is a function $f_{rangehigh}$ which again looks up the deceleration ability table and returns the highest speed which has the same deceleration as its argument. The condition for v_1 to be valid is:

$$f_{rangehigh}(v_2) \geq v_1 \quad (7)$$

If this is not the case, then a stepwise process is used to calculate the valid v_1 . First we calculate the position where the allowed speed = $v_2' = f_{rangehigh}(v_2)$ using the basic brake distance formula (2) above:

$$pos_{v_2'} = pos_p - \left(\frac{v_2'^2}{2 \cdot dec} - \frac{v_2^2}{2 \cdot dec} \right) \quad (8)$$

where dec is calculated according to formula (6) above. We now know the allowed speed at position $pos_{v_2'}$. We can then do a new tentative calculation of v_1 , again using equation (4) but now substituting v_2 by v_2' and pos_p by $pos_{v_2'}$. The new tentative v_1 is then compared with the new (higher) $f_{rangehigh}(v_2')$. The process is continued until we reach a value v_1 which is $\leq f_{rangehigh}(v_2')$. Once the valid v_1 is calculated it shall be compared to the target speed at the new pos_{p-1} , if there is one. The lowest value shall be regarded as the allowed speed at pos_{p-1} . The process continues towards the train, and is interrupted when the end of the brake delay distance is reached.

3.5 Step 5: to interrupt the speed calculation when the end of the brake delay distance is reached

Since the braking ability is zero over the brake delay distance, the calculation of allowed speed shall be interrupted when the end of the brake delay distance is reached, which happens at position $pos_{enddelay}$ as defined in equation (1) above. This means that if $pos_{p-1} < pos_{enddelay}$, then $pos_{enddelay}$ shall be used instead of pos_{p-1} in the last calculation according to step 4. When the allowed speed at $pos_{enddelay}$ is calculated, this will not be updated anymore, except that it is replaced if there is a target speed which is lower (between the train and $pos_{enddelay}$).

4 Additions to the algorithm to be able to handle various special ETCS requirements

In chapter 3, a basic algorithm to calculate the highest possible speed a train can run while still being able to obey the speed restrictions in the restriction table and considering the impact of gradients (the gradient table) and speed (the brake ability table) to the trains deceleration. In ETCS, and any other ATP system, there are many other requirements which need to be satisfied, e.g.

- The system shall be able to calculate allowed speed both for service brake and for emergency brake, the emergency brake being a minimum delay backup, should the service brake fail



- Various speeds for informational purposes, such as permitted speed P and warning speed W shall also be possible to calculate
- Targets may have speed margins for different purposes (e.g. for service brake intervention and emergency brake intervention), and the supervision shall be interrupted when the target speed + speed margin is reached.
- A stop signal position ahead of the train may have a release speed associated with it, in order to make it possible for the train to reach the position where new signal information (possibly clear) becomes available. This is the case when transponders are used to convey the signalling information from the track to the train.

4.1 Allowed speed for service brake and for emergency brake

Service brake and emergency brake have different brake delays and their own deceleration ability tables. The allowed speed to avoid service brake and emergency brake can be calculated by running the above algorithm twice, once with the service brake delay and deceleration table and once with the emergency brake deceleration table. Another method which requires fewer calculations would be to first calculate the allowed speed to avoid emergency brake and then disable all restrictions except the one which was found to be most restrictive for the emergency brake before the algorithm is reused again for the service brake. If however, there is a stop signal among the targets, then this should always be included in the service brake calculations even if it is not the most restrictive emergency brake target, since such a target may be positioned significantly closer for service brake calculations than for emergency brake calculations (at least in ETCS). This is necessary in order to make sure that the train stops before the stop signal even when there is a large safety distance behind the signal.

4.2 Various speeds for informational purposes

In ETCS, the driver has to breach two speed limits before an automatic service brake is issued. The first limit is the permitted speed P, which is showed to the driver during normal operation. As long as the driver runs the train below the P speed, he or she will have at least 5 second margin before ETCS would issue an automatic brake. If the driver exceeds the P speed with a certain margin, the ETCS system will issue a visible and audible warning to the driver. When ETCS issues the warning, the driver has still 3 seconds to react and start braking, before ETCS issues an automatic brake. Since W and P are defined in delay time units, they can be calculated by adding the W and P margins to the brake delay time before running the algorithm. To do this, two extra fictive delay end positions are calculated, one for W and one for P:

$$pos_{enddelayw} = pos_{enddelay} + 3s \cdot v$$

$$pos_{enddelayp} = pos_{enddelayw} + 2s \cdot v$$

where v is the same speed as discussed in chapter 3.3. The calculations are carried out in one pass, but they are interrupted at different positions for the



calculation of W and P than for the allowed speed to avoid service brake. This will cause W to be lower than the “allowed speed” and P to be lower than W, and it results in the desired time margins for the driver to react.

4.3 Speed margins and release speeds

For a signal speed or static speed restriction, there is a margin from the nominal speed to the speed when automatic service brake intervention is issued. There is also a margin between service brake intervention and emergency brake intervention. Unnecessary service brake intervention is thereby avoided as long as the driver drives close to the nominal speed, and unnecessary emergency brake is avoided when service brake is sufficient to do the job.

When the train approaches a restriction which has a margin, it is not necessary to brake the train down to the nominal speed – it is enough to brake it down to the nominal speed plus margin. Release speeds are similar to speed margins – the release speed is a speed margin above zero which enabled the train to approach the transponder close to the main signal.

Speed margins and release speeds are handled by calculating a position before the actual target where the allowed speed is equal to the target speed plus margin (or equal to release speed). The margin targets are used instead of the real targets in the calculations.

5 Conclusion

The ETCS requirements that gradients and targets shall be independently separated, and that the trains deceleration shall be defined in tables related to speed introduces considerable complexity in the supervision algorithms needed. This article suggests a method of combining the gradient and target tables, in order to master this complexity and to limit the necessary number of square root calculations. The allowed speed is calculated from the most distant target and backwards towards the position of the train. Fictive train delay times are used to produce the different speeds that are required for informational purposes.

References

- [1] Alcatel, Alstom, ..., ERTMS/ETCS Class 1 System Requirements Specification, Chapter 7 ERTMS/ETCS language. SUBSET-026-7, Issue 2.2.2, Date 020201.
- [2] Alcatel, Alstom, ..., ERTMS/ETCS Class 1 System Requirements Specification, FFFIS STM Application Layer. SUBSET-058, Issue 2.1.1, Date 19-11-03.



This page intentionally left blank

ICONIS: the window for URBALIS controlled automatic METRO

P. Noury

ALSTOM Transport Information Solutions, France

Abstract

ICONIS implements the concept of an Integrated Control Centre (ICC) and federates former legacy systems such as Automatic Train Supervision, Supervisory Control and Data Acquisition, Passenger Information System, Public Address system and Video-Surveillance. ICONIS is the ultimate solution to craft operator missions according to operational considerations and forgetting about technology.

It is built on the latest development of computer technology and available hardware and software for delivered systems lasting over generations of underlying software.

ICONIS SCADA server monitors all the systems which ensure trains' operations – Interlocking, Automatic Train Control, Automatic Train Protection – and it provides the operator with a live dashboard on system statuses to improve operational availability.

ICONIS incorporates advanced functions for automatic metro to enhance passenger safety. A function monitors platforms to detect passengers on the track and interacts with ATC and video-surveillance systems avoiding high investment in platform screen doors.

Furthermore, ICONIS provides a very efficient operator interface with the Time Distance Graph graphical function and supports operator decisions via Conflict Detection and Solving function.

ICONIS takes care of the renovation project and offers a very modular approach in order to fit exactly to the function to be upgraded. It builds a new base that is suited for further gradual renovation of the other functions aiming at a fully Integrated Control Centre. Whatever the metro system, ICONIS is the solution with the support of URBALIS.

Keywords: integrated control centre.



1 Introduction

ALSTOM's ICONIS™ implements the concept of Integrated Control Centre (ICC) as part of the URBALIS solution to monitor and control transit rail applications.

URBALIS™ is the advanced signalling solutions for all metro and transit rail applications. This solution uses Communication Based Train Control (CBTC) and addresses driver and driverless projects.

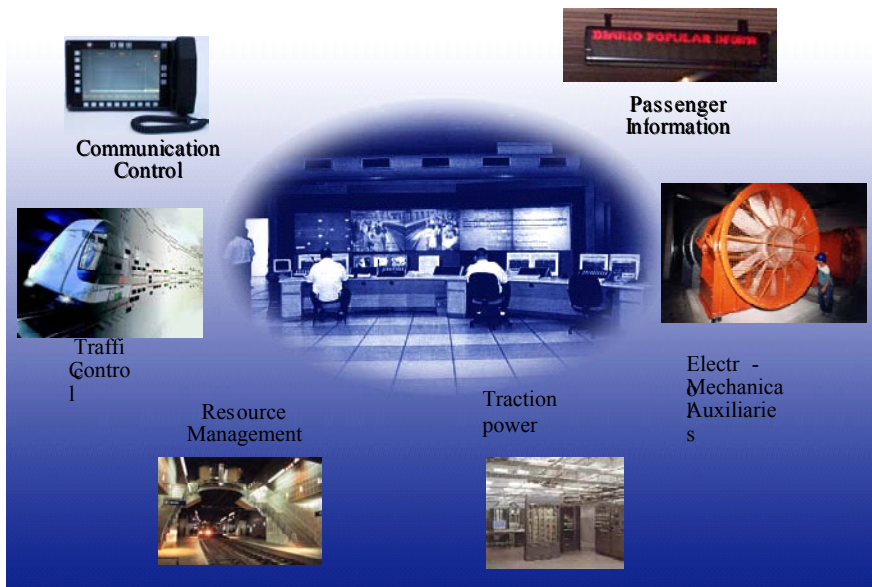


Figure 1: Integrated control centre.

ICONIS™ federates into a single system former legacy system such as Automatic Train Supervision (ATS) system, Supervisory and Data Acquisition system (SCADA), Passenger Information system (PIS), Public address system (PA), Video Surveillance (CCTV).

This concept provides a smart interface to the control room operator who is able to make the proper decisions in order to run efficiently the metro system and deliver the service required by the customer.

In this sense ICONIS is the single window over the URBALIS solution to monitor and control automatic metros. Through this window, information of different sources and natures may be aggregated together to display the current situation in one single screen.

2 ICONIS building blocks

ICONIS provides a single Human to Machine Interface, based on highly flexible graphic mimic diagrams. It also provides a common alarm management system



in order to build a consistent historian of events for on-line analysis and post mortem investigations.

It is both modular and scalable so that it may only implement one or few functions up to the complete package. It is based on 5 basic products suites:

- S2K which encompasses the basic mechanisms for monitoring and control
- Automatic Train Supervision functions
- Field level equipment based on market available software and hardware
- Engineering tools for data base and HMI configuration and simulation tools for testing and training
- Infra ICC for communication management and decision support

These 5 functions may be assembled according to different standard configurations and further called:

- ATS (Supervision of Traffic) when most of the functions are addressing train management,
- SCADA (Supervision of infrastructure) when most of the functions are addressing the management of the infrastructure.
- ICC (Integrated Control Centre) when the above are covered and additional functions are implemented to manage incidents, passengers, resources such as rolling stocks or crews.

When additional functions are needed, they can be added easily to ICONIS standard platform thanks to the use of Window underlying technologies.

This architecture illustrates the concept of software back plane and provides for the unique advantage that new functions do not require re-testing the existing ones. All the functions are fully independent and solely interfaced with the database.

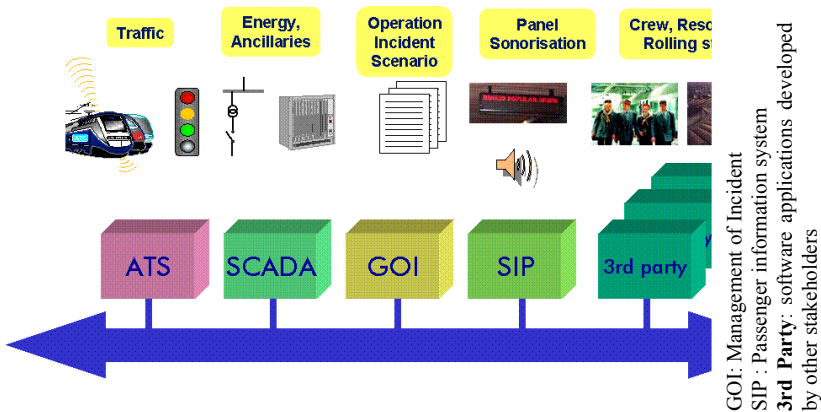


Figure 2: ICONIS ICC software architecture.

3 Basic technology choices

The ALSTOM assumptions are based on the availability on the market of a wide range of computers with multiple processors capability together with operating



systems providing a full library of functions. The drawback of these features is the rapid change of the technology much faster than the evolution of the railway market. Then ICONIS is based on open technologies allowing underneath hardware and middleware to evolve without questioning the application functions and features.

The architecture selected is the classical 3-tiers well known in information technology. The advantages are well known, but specifically for railway application, it provides:

- Applications independence
- Client access to all the application data whatever they are located
- Consistent integration as data are modelled through object classes known across the system
- Server redundancy
- Multiple simultaneous clients

Furthermore the field information is represented along with a common database compliant to OPC (OPEN Connectivity). This database supports the representation of each and any railway device. OPC is a well known developed data base very extensively used in industry and supported by multiple software editors (see www.OPCfoundation.org).

This has led to the concept of software components. A component is a software piece implementing an elementary function of the ATS and SCADA domain. The component is interfaced with the database only.

Each component is fully validated and tested and can be considered as an off the self software.

The components may be of very different nature:

- Driver that adapts an external communication protocol to the data base
- The alarm management function
- The signalling rules verification function
- The reporting function

In the other end the concept of software backplane has been implemented. It encapsulates the object oriented infrastructure and the OPC database. It provides all the resources expected by the software components in a common way.

Building an application becomes the art of selecting the right set of component to fulfil client requirements and plugging them into the software backplane. A complete application is then implemented without writing a single line of code.

4 Communication

ICONIS servers are permanently connected to the URBALIS secure controllers by means of IP communication links.

To support the communication, the devices use the client servers approach but not in a straightforward implementation in order to ensure:

- Independence between the URBALIS network and the ICONIS network
- Make the URBALIS secure controller non sensitive to the number of ICONIS servers



This implies the implementation of Proxy devices that will behave as client to URBALIS controllers and servers to ICONIS clients. These proxies will also filter the traffic and avoid that the ICONIS intra traffic that carries large information chunks may perturb the cyclic traffic of the URBALIS controllers.

There will be two proxy devices each one looking at one IP net of the URBALIS communication system. Each proxy will be connected to each redundant ICONIS server in order to provide the best availability.

This organisation provides both:

- Seamless integration as the messages are the same along the link
- Flexibility of the architecture to answer the range of user requirements
- Performance, as each network can be used to its optimum load without having to care about the load and QoS (Quality of Service) of the other one.

5 Monitoring the environment

5.1 System monitoring

ICONIS, as mentioned before, is able to support a SCADA system with the same underlying technology as the ATS. The SCADA servers exhibit the same interface as the ATS servers so that the operator workstations may access any information managed by either server in a transparent way.

Then the SCADA functions for monitoring and controlling the infrastructure can be applied to the system itself. These capabilities provide the operator with a dashboard summarising the system statuses and telling him if the system is able to achieve the required tasks or if it needs service and where.

5.2 Passengers monitoring

In the field of automatic metro, it is of essence to ensure the passenger security. In addition to the legacy domains, it is necessary to improve the passenger security at the transfer level for boarding and un-boarding the trains.

Besides the fact that the SCADA is able to manage platform screen doors there is a need for lighter system to perform similar functions in application such as outdoor stations, refurbishment of existing stations or simply to minimise investment.

Platforms will be equipped with dual detection systems capable of detecting a person falling on the track and reporting on time to the Automatic Train Control system. Those systems should be carefully monitored to ensure that proper mitigation is set in case of an incident and that no false alarm occur and perturb the traffic.

This is achieved by a close cooperation between the detection system, the CCTV system and the SCADA Incident Management Function. The principle is that, when an incident is detected, the alarm is sent to both URBALIS controllers and the ICONIS ICC. If a train is approaching, URBALIS will stop the train, if the train is far from approaching then the information is analysed by the operator



via auxiliary cameras and Public Address system. This may give time for the person to step back to platform or to ask personnel close by to come and rescue the person.

To help the operator the ICONIS ICC will present him the situation in context (where are the trains, presence of personnel at this location...) together with a checklist for launching the mitigation actions. When incident is mitigated, the ICC system will propose a report of it for filing the case and feeding the database for improving the incident procedures.

This illustrates the fact that even if the functions for passenger monitoring are typically under the SCADA scope, their management is the responsibility of the traffic operator principally dealing with ATS.

In conclusion ICONIS breaks the technology barriers between ATS and SCADA and supports the definition of operator mission only based on functional aspects of the train system operations.

6 ICONIS train functions

In order to ensure automatic train operation and traffic regulation for URBALIS controlled metro systems, the ICONIS ATS system addresses the following traditional main functions:

- Field equipment management,
- Train describer,
- Train identification,
- On-line timetable management,
- Automatic route setting,
- Automatic train regulation,
- Depot management,
- Support functions (e.g. Alarms & Events management),
- Off-line functions.

These functions and the data flow between them are illustrated in the following figure.

The Human Machine Interface offers to each of the above mentioned functions, dedicated interfaces which allow the operator to perform the required controls and information monitoring.

All the above listed functions rely on the integration of the ATS with other specialised subsystems and principally the Interlocking and the Automatic Train Control subsystems. The following is set to illustrate the seamless integration between URBALIS and ICONIS.

The Automatic Train Control management function communicates directly with track-side and embedded URBALIS sub-systems in order to:

- Report the train location and the train status whatever the trains on the line situation,
- Report train emergencies thanks to fire alarm detections or passengers emergency handles; information is used to query the relevant camera in order to visually confirm the issue and trigger dialogue with passengers,
- Report trackside emergencies thanks to tunnel fire alarm detection, platform monitoring system, platform plungers,
- Collect maintenance information from the on-board and track side equipment concerning principally signalling, ATC, Rolling Stock



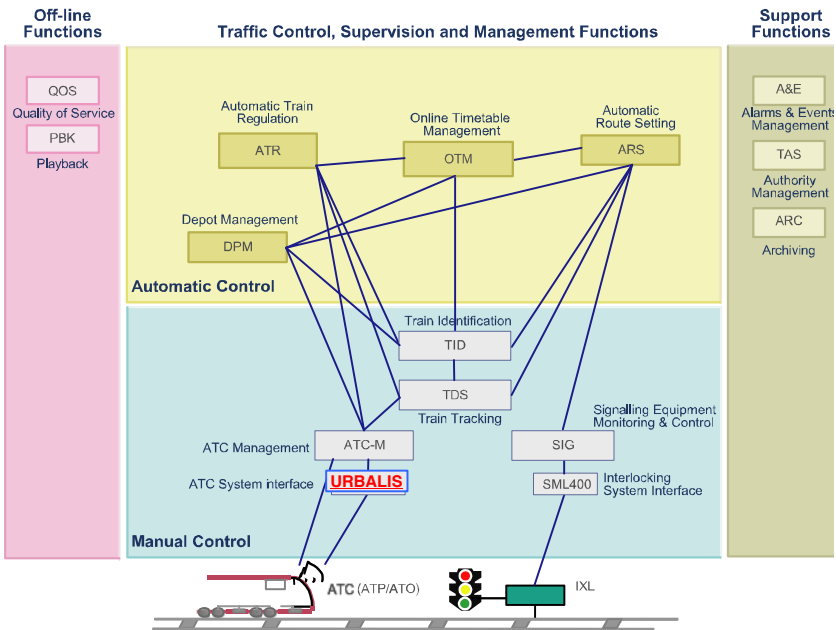


Figure 3.

To achieve its mission, ICONIS implements secure commands when it deals with safety related actions: Temporary Speed Restriction (TSR), Geographical Automatic Mode Authorization (GAMA), Individual Train Automatic Mode Authorization (ITAMA), Track-circuit in/out operation, Initial/Train evacuation reset, Line/siding switching.

Furthermore the URBALIS Management function is in charge of handling information for detailed train operation:

- Train readiness modes,
- Skip-Stop station,
- Departure Time and Arrival Time,
- Hold facilities
- Energy saving,
- Change of ends,
- Speed restriction for un-monitored zone crossing
- Connect & disconnect carriage pairs at depots or sidings.

(Train, Platform, General Platforms)

The main function of ICONIS ATS is to achieve this, the Train Descriptor concept has been introduced, which consists in processing trains locations in order to detect relevant events for ATS operation and display trains' movements on HMI.

The train detection movement is based on moving block data provided cyclically by URBALIS ATC system (train identity and location).

As a train progresses along the line, the detection of significant events (e.g. trigger crossing, train arrival, etc) is achieved. These events are sent to the other ATS functions (e.g. Automatic Train Regulation, Automatic Route Setting).



The operators can follow on HMI the train's progression thanks to the train indicators. Each train indicator is associated to a given portion of track and indicates the identity of the train located within.

7 Time distance graph and conflict solving

7.1 Time distance graphs

Time Distance Graph (TDG) is a software component, within the range of ICONIS ATS components, intended to help users to monitor and regulate the train traffic circulation.

Connected to the other functionalities of ICONIS ATS (e.g. Timetable Management, Estimated Time of Arrival, Infrastructure Management, etc.), It gives a clear graphical view of the train circulation, infrastructure state, abnormalities and detected conflicts.

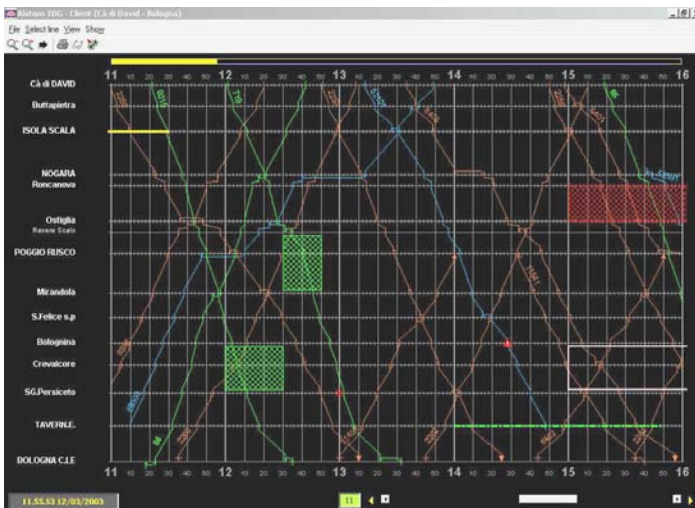


Figure 4.

TDG can be used either in online mode or in off-line mode, in order to plan in advance the train circulation taking account the situation of the infrastructure or temporary modifications to timetables. It is easy to use and highly configurable. It allows the operator to see in advance problems on train circulation in a clear and intuitive way, then to act quickly to solve the problems.

TDG displays:

- Infrastructure provisions (speed restriction or way restriction) by rectangles covering the area involved in the infrastructure provision.
- Abnormalities (general problems that can create any disturbance on the traffic circulation).



- Conflicts between trains or between a train and the infrastructure by means of an appropriate symbol.

TDG is a convenient summary tool to manage the traffic in a smooth and efficient way

7.2 Conflict detection and solving

Each train has its own resources allocated such as infrastructure, rolling stock, crew, etc. These resources are based on daily and current timetable requirements. When disruptions create conflicts in traffic, Conflict Detection and Solving (CDS) checks the resource allocations availability or incompatible usage of resources which lead to conflicts. Then CDS suggests resource allocation (when and what) reassignment and trip modifications to the dispatcher operator to solve these conflicts.

Even if a conflict can be always described by an occupancy resource conflict, it is possible to list four kinds of conflicts:

- Headway conflicts: between two stations, no flanked ways are possible and trains are running in the same direction. Headway time is not respected and trains require the block section line at the same time,
- Cross conflicts: between two stations, one block section line and trains are running in the opposite directions. Trains needs the block section line at the same time,
- Route conflicts: in the station. Trains require two incompatible routes at the same time,
- Platform conflicts: in the station. Trains require the same platform at the same time.

The solution is determined for some (priority mode) or all conflicts (objective mode) detected inside a solving window included in the detection time window.

Each time a conflict is detected, CDS immediately solves it and provides the operator, through either the Time Distance Graph or the Tabular Time Table, with:

- The list of all the conflicts and possible solutions,
- The new proposal trip plan that solves the conflicts.

The solution of a conflict is obtained:

- By modifying the trip of the involved trains in order to avoid the conflict,
- Or selecting alternative resources satisfying the constraints.

8 Common functions

ICONIS ICC thanks to its structure is able to collect all the events that occur into the system: Archiving, Playback and Training facilities can be used by operators.



8.1 Archiving

The Archiving function includes two main activities:

- Archiving: storage of data and backup of these data on an external support.
- Restore: restore of the data from the external support.

8.2 Playback

The Playback function allows the data related to a period of operation to be recorded and subsequently replayed. Playback helps in analyzing specific incidents by reviewing the recorded system data (events, alarms, etc).

8.3 Training facilities

The training on the ATS functions is feasible on the ATS itself. To achieve this, the ATS is connected to an URBALIS simulator. This simulator will behave as the true system and incorporate train behaviour. Then an instructor station may inject scenario into the simulator for the operator to face real situations and to be trained to make proper decision in time.

9 Conclusions

As shown above URBALIS and ICONIS are two sides of the ALSTOM knowledge in transit system control and operations.

There deliberate structure merges all the legacy functions of transit train system together in order to define each operator function based on operational aspects without any technological constraints.

The traditional separation between Train movement, Lateral signalling and infrastructure management has disappeared, but without compromise on security or availability.

New functions have been added to legacy systems such as security of passengers on platform, management of incidents, selective CCTV to facilitate operations

The integration of function including service functions allow to building a single historian database. That database would represent the full history of the transit system and support comprehensive replay of any operation event for understanding incident cause and development, storing the information for authorities investigation, improving the operational rule to minimise the consequences of an incident.

The technology choices provide for a long life of applications even if the underlying computer technology evolves rapidly. In addition the openness of the database (OPC and SQL server) allow for development of specific functions while the system is already up and running.



Automatic train operation system for the high speed Shinkansen train

Y. Yasui

*Advanced Railway System Development Center,
East Japan Railway Company, Japan*

Abstract

In train operation over 300 km/h, drivers are supposed to operate the handle for acceleration or deceleration quite often due to speed restriction at the curve. To ensure on-schedule operation and lighten the burden imposed on drivers under high speed operation, we developed the automatic train operation system for the Shinkansen train. This system automatically controls the speed to follow the target speed of operation as well as obey Automatic Train Control (ATC). The target speed is set by taking account of the ATC restriction, on-schedule running, and energy-efficient operation. The system was applied to running tests using a series E2 Shinkansen train from Morioka to Hachinohe, about 100 km, on Tohoku Shinkansen line. The test results tell that the accuracy of on-schedule operation is 4 s longer than the simulated running time, and the accuracy of following the target speed is within 2 km/h at the maximum speed of 320 km/h. We conclude that the system performance is satisfactory for the secure speed control and on-schedule operation.

Keywords: automatic train operation, automatic train control.

1 Introduction

East Japan Railway Company is developing the Shinkansen train with the concepts of high speed, safety, stability, environmental compatibility, and comfortableness. As for high speed, it plans to operate the train at the maximum speed of 360 km/h. To ensure on-schedule operation and lighten the burden imposed on drivers under the high speed operation, we developed the automatic train operation system for the Shinkansen train and carried out the running test on Tohoku Shinkansen line.



2 System outline

Automatic Train Control system (ATC) helps to keep train intervals fixed by braking the train, but in the high speed Shinkansen it is necessary to control the train in the shortest running time by powering and braking. This system automatically controls the speed by powering and braking to follow the target speed of operation as well as obey the ATC. The target speed is set by taking account of the ATC restriction, on-schedule running, and energy-efficient operation. The basic action of the system is as follows. Firstly, it accelerates up to the speed just below the ATC restriction. Secondly, it keeps the train at the maximum permissible speed with extreme accuracy. Thirdly, it stops accelerating at the point allowing on-schedule operation to the destination as it runs. The accurate operation at the maximum permissible speed reduces the loss of running time and produces the marginal time to the scheduled running time. According to the marginal time, it stops accelerating to save energy consumption, fig. 1.

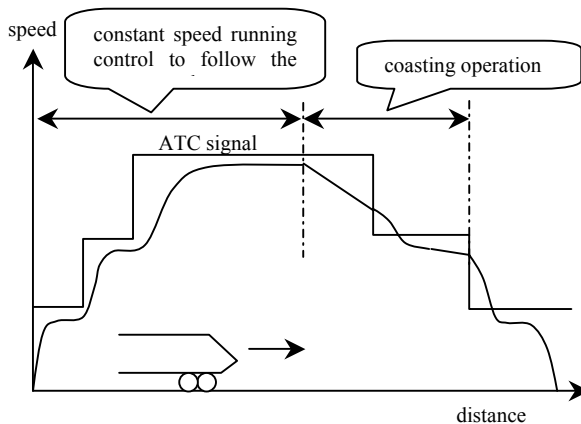


Figure 1: Running pattern image.

3 System constitution

This system consists of a switch in the drivers cab and the operation system connected to some devices, fig. 2. The switch in the drivers cab is pushed to start the automatic train operation. The operation system is connected to monitoring equipment in order to get information of kilometre and ATC signal aspect by way of ATC equipment, so it recognizes the train location and the restriction speed. It is also connected to a tachometer generator to get information of the train speed, so it controls notch for powering and braking to follow the target speed. The notch means steps for powering and braking.



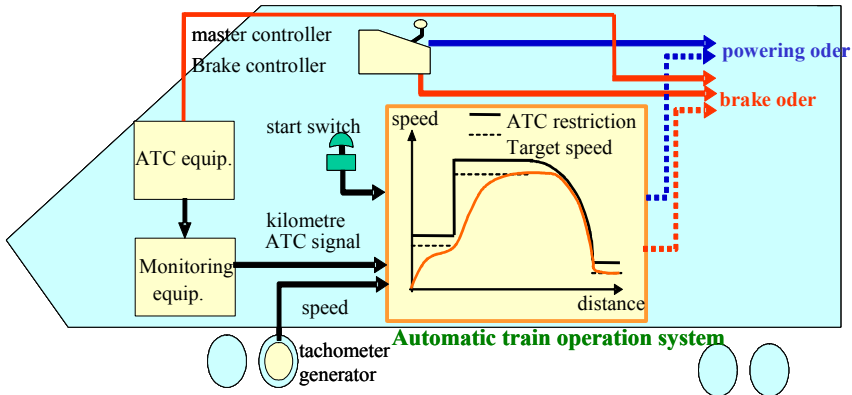


Figure 2: Constitution of automatic train operation system.

4 Function

4.1 Target speed

The system receives the ATC signal aspect from the monitoring equipment and sets the target speed of operation as the speed below the ATC signal aspect by 3km/h. Using information of kilometre, it can set the target speed according to the train location.

4.2 Constant speed running control to follow the target speed

The system controls to keep the speed constant to follow the target speed of operation until the target speed changes. The accuracy of constant speed control is set by within 2 km/h over or below the target speed. Concretely, it orders notch control to absorb changes of the train acceleration or deceleration by gradient resistance and running resistance, and keep riding comfort as it is. For this control, it has data of notch choice set by the train location and running speed zone and calculates relevant notch forward and back according to the changes of train acceleration or deceleration.

4.3 On-schedule operation

The system counts the time passed from leaving a station and calculates the difference between running time left and the target running time. When the difference is plus, it stops accelerating to shift to coasting operation for punctuality and energy saving. The accuracy of on-schedule operation is set within 10 seconds more or less than the target running time. The basic control routine for coasting and efficient operation is as follows, fig. 3:

- (1) In the fastest operation pattern, the time to arrive at a next station by coasting from a point is defined as T_{ex} . The system has T_{ex} data by simulation in advance.



- (2) The system calculates the time passed from leaving a station, T_{sx} . It compares a sum of T_{sx} and T_{ex} with scheduled time, T_t . At a point where the sum of T_{sx} and T_{ex} is equal to T_t , it starts coasting operation.
- (3) In case of driver's manual control, it cancels the operation once and start again coasting operation when the condition is satisfied.

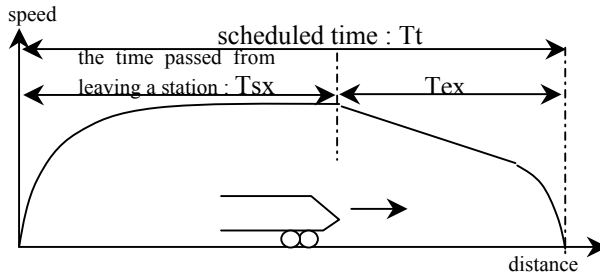


Figure 3: On-schedule operation image.

5 Test result

The system was applied to running tests using series E2 Shinkansen train from Morioka to Hachinohe, about 100km, on Tohoku Shinkansen line.

5.1 Constant speed running control to follow the target speed

The accuracy to follow the target speed was within 2 km/h in almost all section of running tests. There were two points where the accuracy was 2.2 km/h over or below the target speed. It is because of gradient fluctuation such as from plus 9 to minus 8, so it is improved by adjustment of control parameters of notch choice data.

5.2 On-schedule operation

Firstly, the fastest operation under the ATC restriction was tested. Secondly, we set the marginal time as 30 s. On-schedule operation was tested with the target running time as a sum of the running time of the fastest operation and 30 s. As table 1 shows, the running time of the fastest operation was 1591 s. The target running time was set as 1621 s. The running time of 30 s marginal time operation was 1625 s, so the difference between the target and result was 4 s accuracy. Figure 4 shows the result of running test with marginal time of 30 s.

6 Evaluation

6.1 Constant speed running control to follow the target speed

It is a satisfactory result that the accuracy of constant speed running control was within 2 km/h over or below the target speed in all section except for abrupt gradient fluctuation. The system performed stable control to absorb gradient resistance and running resistance from lower to higher speed zone, so it helps to



lighten operating load of drivers, who are supposed to operate the handle for acceleration or deceleration quite often, and realize constant patterned operation in all time.

Table 1: The accuracy of on-schedule operation and energy consumption (running section: Morioka to Hachinohe).

	The fastest operation	Marginal time 30 sec.
Target running time (second)	-	1621
Running time (second)	1591	1625
Powering energy (kWh)	2443	2173
Regenerative braking energy (kWh)	353	230
Consumption energy (kWh)	2090	1943

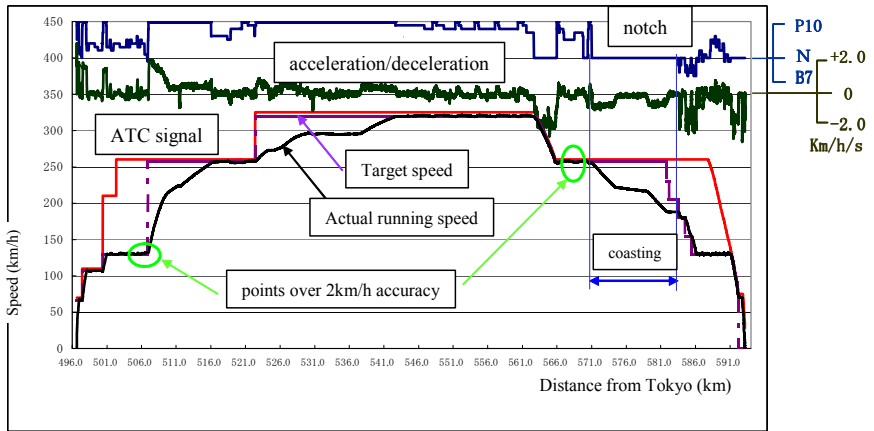


Figure 4: The result of running test from Morioka to Hachinohe.

6.2 On-schedule operation

On-schedule operation was realized by shifting accelerating to coasting according to marginal time of running. The accuracy of punctuality was 4 s, which means that the simulation data about target running time was highly accurate. It satisfies what we set at a target, and we regard the on-schedule operation as possible in the high speed Shinkansen.

7 Conclusion

It is concluded that the system performance is satisfactory for secure speed control, constant speed running control, and on-schedule operation. The system also realizes energy saving operation. For the practical use, we will accumulate

the data of various running patterns in the high speed Shinkansen test train aiming at 400 km/h running.

References

- [1] Operation theory research organization, Operation theory, 2002. (Japanese)
- [2] Hajime Ochiai, Masahiko Mitsukawa, Hiromichi Nishio, Takeo Yoshimoto, Yoshitaka Yasui, Development of operation assistant system, Cybernetics, 2003. (Japanese)



Section 8

Train location

This page intentionally left blank

Design of experiment for the validation of ATP/ATC odometry algorithms

M. Malvezzi¹, G. Cocci² & A. Tarasconi²

¹*Department of Energetics S. Stecco, University of Florence, Firenze, Italy*

²*Trenitalia S.p.A., Firenze, Italy*

Abstract

The safety and efficiency of a railway control systems depend on many factors, one of these being the precision of train position and speed estimation. The knowledge of the estimation precision is important in the developing phase of the procedure and in the integration with the automatic train protection system, in order to properly assess the safety margins used to define the system intervention. The estimation is performed by the so-called odometry system, which includes a series of sensors and a software procedure. The data set used to test the algorithm has to represent all the real conditions that will be met in normal conditions. The criteria used to define the set of tests is based on the DOE (Design Of Experiment) technique, which allows one to obtain a compromise between the number of needed tests and the complete reproduction of possible cases. Once the set of tests are defined, they can be reproduced on a test rig, like the MI6 rig, developed and realized by Trenitalia. The parameters that mainly influence the algorithm precision are the wheel/rail adhesion, the type of train, the type of operation and the gradient of the line. For each parameter a series of classes has been defined and combined in order to obtain a set of tests able to represent the behaviour of the algorithm in all the cases it could meet in reality and then to obtain a reliable estimation of its prevision. In this work the criteria used to define the data set will be described. Then this set will be used to test the odometry algorithm developed for the SCMT, the Italian ATP system and the results will be presented.



1 Introduction

An Automatic Train Control system provides a high level of train safety, allowing one to increase the running speed and the line capacity. These systems automatically perform their monitoring and control action in order to prevent against the consequences of driver errors. Usually an ATP system includes three components:

- the *wayside equipment*, which generates codes to be transmitted to the train;
- the *track-to-train* transmission system, which transmits information from the wayside equipment to the train;
- the *train on board equipment*, which elaborates the information received from the track and from sensors mounted on the train and decides the actions to be performed.

The track subsystem gives to the train a series of information, for instance:

- the current position and the line gradient;
- the distance to targets (point in correspondence of which speed restrictions have to be achieved);
- the target speed (i.e. the speed that has not to be exceeded when the train passes through a target point).

This information is communicated to the train with the aid of fixed balises or another kind of absolute information. Between two consecutive information points on the line, the on-board subsystem perform a dead reckoning of the speed and the travelled distance and calculates:

- the minimum distance that allows to respect the speed restrictions at the next objective points: this value depends on the actual train speed, on the braking parameters and on the objective speed;
- the distance to the next information points.

If the difference between the distance to one or more of the next objective points and the distance that allows to respect the speed restriction is smaller than a fixed value, the on-board subsystem intervenes, for instance by activating the emergency braking.

A correct estimate of distance to target and actual velocity is then crucial to evaluate residual braking resources, in terms of available deceleration, in order to reach the targets at the required speeds.

In order to perform this estimate, some of the ERTMS implementations (European Railways Traffic Management System, the system under development for the European railways) uses a set of sensors including: encoders positioned on two independent wheels, a radar sensor positioned on the first vehicle case, and a longitudinal accelerometer.



The performance of dead reckoning by means of odometry (in terms of precision in the estimation of speed and travelled distance) depends on how the information from different types of sensors is fused. For example, the estimation obtained with the measure of one or more wheel speeds may fail if degraded adhesion in the wheel-rail interaction occurs, due to rain, fog, ice, leaves, and so on, and the train is accelerating or braking, i.e. when pure rolling conditions between the wheel and the rail do not hold anymore, and macroscopic sliding occurs on one or more of the axles equipped with odometry sensors. The accelerometer sensor is sensitive to the inclination with respect to the horizontal direction, so it can give erroneous information in presence of a line gradient or pitch motion of the vehicle where it is mounted.

2 The odometry system

This module named *odometry system* evaluates the speed and train position during its run. It includes a series of sensors (for example wheel angular speed sensors, radar Doppler, longitudinal accelerometer) and a procedure (named “Odometry Algorithm”), that *fuses* data from the track and from the sensors mounted on the train in order to perform the *odometric estimation*. The values calculated by the odometric system are communicated to the ATP on board module, that compares them with the braking curves and decides an eventual intervention.

The odometry system performance, in terms of accuracy of speed and distance estimation, may affect the whole ATP system performance, in particular:

- if the odometry system tends to under-estimate the speed and the current position, the *safety* of the ATP system may be affected, since the train “thinks” to have a speed lower than the real one and/or to be at a distance from the objective greater than the real one, then a possible intervention of the control system could be delayed and the objective (for example stopping at a red signal) could be not respected;
- on the other hand, if the odometry system tends to over-estimate the speed and the current position, the *efficiency* of the ATP system may be affected, in this case the train “thinks” to have a speed greater than the real one and/or to be at a distance from the objective shorter than the real one, then a possible intervention of the control system or could be not necessary or anticipated.

For these reasons a particular care have to be put in the design and testing of such devices.

3 Sensors used to perform the odometric estimation

3.1 Wheel speed sensors

The sensors that measure the wheel angular speed are widely used in the on board subsystems (for example WSP and anti slip devices). Parameters limiting



the accuracy of this kind of measure are the resolution of the sensor, the sampling time, electrical noise, mechanical imperfections such as backlash and, most importantly, slip and slide between the wheel and track.

Some research has been carried out to find methods for correcting the speed and position error caused by slip and slide using wheel angular speed sensors. The odometry algorithm developed for the SCMT (Sistema di Controllo Marcia Treno), the Italian ATP system, is based on the measures obtained from two wheel angular speed sensors. The algorithm is able to detect when the wheels are sliding and to then to properly correct the speed estimation [5–7].

3.2 Accelerometer

Linear accelerometers can be used to measure the linear acceleration of the vehicle, by integrating the acceleration signals both speed and position data can be derived. For this measurement method the exact initial values of speed and position are required.

There are many types of possible errors that could arise using this type of sensor, in particular, the measure is sensitive to the angular position of the sensor and then to the motion of the body on which it is mounted. For example, if a mono-axial accelerometer is used to measure the longitudinal acceleration of the train, the measure will be affected by the pitch motion of the vehicle and by the line gradient.

3.3 Radar Doppler

This type of sensor measures the relative speed between itself and a surface by detecting the frequency shift of a signal. This phenomenon is known as the Doppler effect. The measure of this kind of sensor may be affected by some possible sources of error:

- very smooth surfaces may cause loss of discrimination in the reflected signal,
- change in the radiation angle caused by the pitch and tilting motion of the body on which the sensor is mounted,
- existence of vibration can create noise in the data,
- existence of radar noise and bias errors.

3.4 Sensor fusion

In order to obtain a reliable odometric estimation and to prevent system failure the use of different types of sensors is necessary. The different measures has to be combined, the system has to identify the conditions in which the sensors are operating and to choose the measure that, in every situation, has the maximum reliability.

The odometry algorithm is then designed to perform an estimation of the *state* of the system (it has to be able to recognize the situations in which one or more of the sensor measures are affected by an error due to the operational conditions)



and to choose for every condition, the estimation procedure that with the greater probability is the more corrected.

The following table summarizes the sensitivity of each sensor with respect to particular operational conditions.

Table 1: Sensor measure sensitivities.

Type of sensor	Accelerometer	Angular speed sensors	Radar Doppler
Sensitivity with respect to			
Adhesion between the wheels and the rail	NO	YES	<i>It is not possible to find a relation between the radar Doppler sensor performance and these conditions</i>
WSP and pneumatic brake dynamics	NO	YES	
WSP and electric brake	NO	YES	
Traction system, traction control (anti slip device)	NO	YES	
Mechanical properties of the vehicle and of the traction system	NO	YES	
Information on line gradient	YES	NO	
Vehicle pitch motion	YES	NO	YES
Vehicle roll - tilting	NO	NO	YES
Reflecting properties of the line	NO	NO	YES
Presence of contaminants (snow, water, dead leaves etc.)	NO	NO	YES

4 Design of experiments

The test of the odometry subsystems could be performed by a series of test runs, but this way is practically difficult for these reasons:

- a great number of tests are necessary in order to properly validate device performances;
- some testing conditions cannot be obtained during a run on the rail;
- testing conditions are difficultly controllable and repeatable;
- the testing activities have to be integrated with the normal rail traffic.

This way is not practically suitable in particular in the design phase of the system, when the estimation strategy has to be optimized. In this phase a great amount of data is necessary in order to explore all the operational scenarios that the system could meet in the reality.

Another testing strategy, that effectively can overcome all these problems, is represented by a proper HIL test rig. The feature of the test rig used to test the ATP on board subsystems is described in [1–3].

If only the odometry algorithm has to be validated, the test rig can be used in a off-line mode, in other words the generation of the test cases can be independent from the validation phase. Then, once the type and the number of tests have been decided, a single data base of tests has to be defined. The availability of this data set allows to speed up the testing phase, to have a useful



tool able to compare the performance of different solutions, to perform easily sensitivity analysis and optimization procedures.

This way is obviously not suitable if the whole on board subsystem is tested, because the intervention of the ATP system modify the dynamics of the test.

The tests used to validate the odometry system have to represent all the operative conditions that the system could meet during its activity. In the tests the situations that could affect the safety of the system (situations in which one or more sensor measures are affected by errors and uncertainties) have to be reproduced, in order to test the safety of the system. In order to define a proper set of test cases, an analysis of parameters that mainly influence odometric estimation performance is necessary.

From the analysis of the sensor features the main parameters that affect system performances were identified:

- train type and composition: it has an effect on the wheel dynamics (in presence of poor adhesion conditions) and on the dynamics of the lead vehicle, this parameter then has an effect on wheel speed sensor, accelerometer and radar Doppler measures;
- brake mode: it has an effect on the dynamics of the brake propagation and then on wheel dynamics (in presence of poor adhesion conditions) affecting the wheel speed sensor measure;
- train length;
- line gradient: this parameter has an effect on accelerometer sensor measures (a component given by the projection of the gravity acceleration on the travelling direction is added to the measure of the accelerometer, this component acts as a disturb on the measure and if it is not properly corrected may leads to high estimation errors);
- adhesion conditions: they have an effect on the wheel dynamics and on the dynamics of the lead vehicle, this parameter then has an effect on wheel speed sensor, accelerometer and radar Doppler measures.

The coefficient used to specify the adhesion conditions is the static friction coefficient, imposed at the beginning of the test. During the test this coefficient can change, due to the wheels sliding and their time history.

For each parameter the main significant categories were identified. By the combination of all these cases the set of tests necessary to completely analyse the algorithm behaviour were defined [1]. Once the tests have been realized, the results can be analysed in order to better understand which parameters have an effective influence on system performance and how these parameters affect the precision of the estimation.

Once the test conditions have been defined, the methods for the result elaboration have to be set. Since the number of data is very high the odometric estimation error was evaluated by means of statistical techniques. For each simulation, the travelled distance is divided in elementary segments with a fixed length (for example 200, 500 and 1000 m) and the difference between the estimated distance and the reference value is calculated. At the end of this task three sets of data are available, the results can be compared with the system specifications.



Table 2: Parameters used to design the tests.

Train type and composition
Locomotive, empty freight train
Locomotive, loaded freight train
Locomotive, passenger train
Coach, passenger train, traction applied to the end of the train
Train length
Minimum value (on the basis of braked weight percentage)
Maximum value
Intermediate value between the preceding ones
Line gradient
0
$\pm 9 \text{ ‰}$
$\pm 21 \text{ ‰}$
$\pm 35 \text{ ‰}$
Adhesion between the wheel and the rail μ
0.09
0.07
0.05
0.03
Type of operation
A: Acceleration from 0 km/h up to 100 km/h with the maximum traction torque available (that depends on the adhesion value), three minutes at constant speed and emergency braking
B: Emergency braking from a speed included in the range between the maximum speed and 20 km/h
C: Maximum service braking from a speed included in the range between the maximum speed and 20 km/h
D: Minimum service braking from a speed included in the range between the maximum speed and 20 km/h
E: Acceleration from 0 up to the maximum speed with the 25, 50, and 75% of the available traction torque
F: Maximum/minimum service braking from the maximum allowed speed up to 50% of the maximum speed, followed by an acceleration with the 25, 50, 75, 100% of the available traction torque
G: Acceleration with the 25, 50, 75, 100 % of the available traction torque followed by an emergency braking

5 Examples

The following figures show an example of test. Figure 1 shows the actual and estimated speed and the velocities where the speed sensors are mounted. Figure 2 shows the difference between estimated and actual speed in the test. A single locomotive running on a line with degraded adhesion conditions (reference value 0.05) and no gradient was simulated in this test. The test was carried out using a odometry algorithm that uses two wheel angular speed sensors and a longitudinal accelerometer on the vehicle body. As it can be seen, even if the test conditions are extremely severe, the odometry algorithm error is quite small (the maximum value is about 1.5 m/s). Figure 2 shows that the error on speed estimation increases during a braking phase, with an approximately linear law. This increasing is due to the fact that the accelerometer measure is affected by the



pitch motion of the body where the sensor is mounted (in this case the locomotive). However, during a braking phase the algorithm tends to over-estimate the speed, so the odometer error does not decrease the system safety.

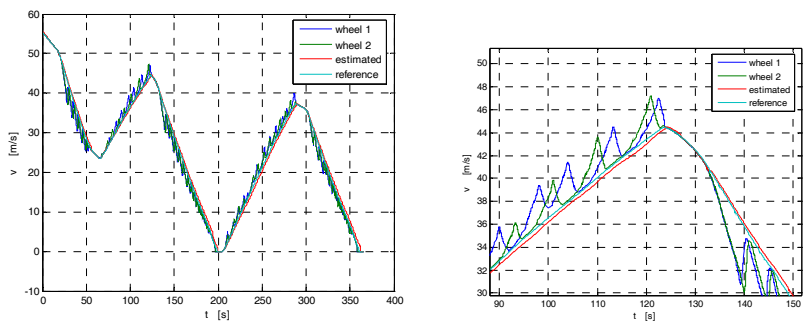


Figure 1: Actual and estimated speed.

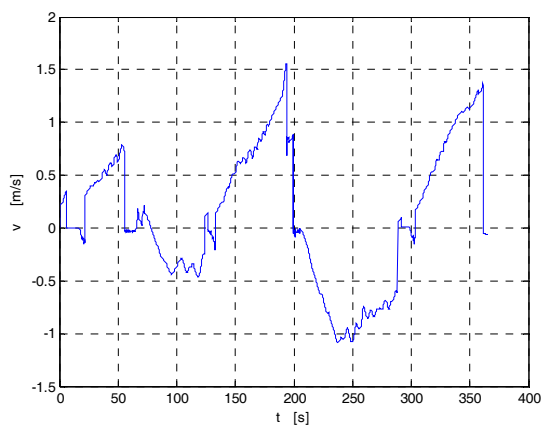


Figure 2: Difference between the estimated and the actual speed.

Figure 3 shows an example of the odometric error distribution, compared with a normal distribution. The example is relative to an algorithm that tends to over estimate the travelled distance, and then the distribution is quite different from the normal one.

6 Conclusions

In the ATP/ATC systems, a correct estimate of distance to target and actual velocity is fundamental to evaluate residual braking resources, in terms of available deceleration, in order to reach the targets at the required speeds. This device has to be properly designed and validated, in order to quantify its performance.

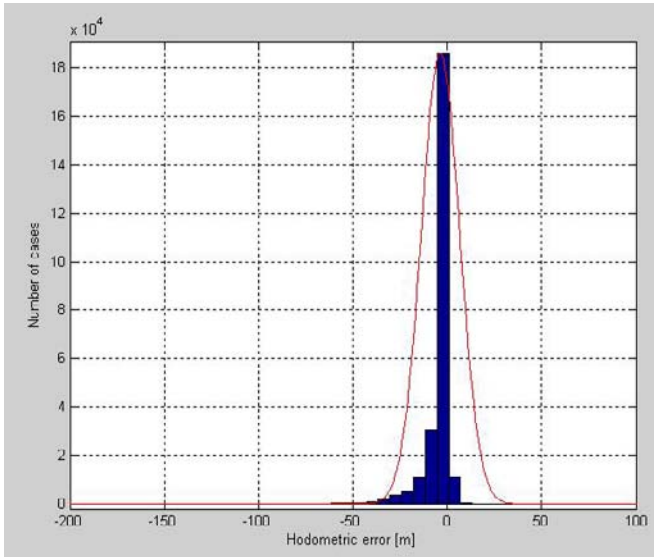


Figure 3: Statistical distribution of the odometric error.

The use of a properly designed HIL test rig can effectively represent an interesting solution for the testing activities necessary to the validation of odometry devices. This test methodology allows to greatly simplify the validation procedure. Test runs on the line are no more necessary and then the costs decreases sensibly. Testing conditions are completely controllable and repeatable this permits for example the comparison between different solutions. Extreme testing conditions (very low adhesion conditions, axles locking up and so on), that are difficult or dangerous to be reproduced on the line, can be easily obtained.

From the analysis of the type of sensors and the algorithm used to perform the odometric estimation the main parameters that may have an influence on system performance were identified. For each parameter a series of classes were defined. The set of data usable for the testing activity was then defined by combining the different classes.

References

- [1] F. Addeo, B. Allotta, M. Malvezzi, L. Pugi, A. Tarasconi, M. Violani, "ATP/ATC subsystem testing and validation using a HIL test rig", Proc. of 9th Int. Conf. on Computer Aided Design, Manufacture and Operation in the Railway and Other Advanced Mass Transit Systems, May 2004, Dresden, Germany.
- [2] P. Toni, B. Allotta, A. Rindi, Un approccio innovativo alla progettazione di banchi prova per sistemi ferroviari, *Ingegneria ferroviaria*, Aprile 2003, n. 4, pp.341-347.



- [3] B. Allotta, M. Malvezzi, L. Pugi, M. Rinchi, A. Rindi, P. Toni, A. Amore, R. Cheli, G. Cocci, P. Presciani, G. Puliatti, "A Test Rig for Evaluating Odometry Algorithms," Proc. of 8th Int. Conf. on Computer Aided Design, Manufacture and Operation in the Railway and Other Advanced Mass Transit Systems, 12 - 14 June 2002, Lemnos, GR.
- [4] TRENITALIA S.p.A. Unità Tecnologie Materiale Rotabile "SCMT Progetto dell'algoritmo per il calcolo della velocità istantanea del treno e dello spazio percorso," UTM/DT.S.PS, 31/10/2000.
- [5] TRENITALIA S.p.A. Unità Tecnologie Materiale Rotabile "Odometria SCMT:Principi generali dell'algoritmo per il calcolo della velocità stimata in caso di pattinamento o slittamento degli assi di misura: Specifica Requisiti Funzionali." Specifica N. 372574 esp. 03, 28/11/2002.
- [6] TRENITALIA S.p.A. Unità Tecnologie Materiale Rotabile "Odometria SCMT:Principi generali dell'algoritmo per il calcolo della velocità stimata in caso di pattinamento o slittamento degli assi di misura: Specifica Requisiti Software" Bozza 4 Specifica N. 373450 esp. 00, 29/11/2002.



Location in railway traffic: generation of a digital map for secure applications

F. Böhringer & A. Geistler
*Institut für Mess- und Regelungstechnik,
University of Karlsruhe, Germany*

Abstract

This paper presents the modelling scheme for a digital map based on the node-link model. To use this map in combination with a sensor system that recognises turnouts by detection and classification of their components, these reference points must be modelled as spatially expanded objects and thus the node-link model must be extended. The digital map consists of the topological and geometrical level. Both levels are modelled using the Extended Entity-Relationship Model and hence easy implementation into a data base is given. The generation of the digital map for a small secondary line demonstrates performance of the successful modelling scheme. A location system for secure applications can be obtained by using an eddy current sensor system in combination with the proposed digital map.

Keywords: digital map, node-link model, Entity-Relationship Model.

1 Introduction

State-of-the-art digital maps are mainly used for non-secure planning functions or in combination with hardware installations along the track such as balises. The node-link model used in digital maps is not able to model reference points such as turnouts as spatially expanded objects [1]. A train-borne sensor system that is able to detect and classify turnouts cannot be used in combination with such maps. The aim of the digital map presented in this contribution is to introduce such reference points and to use classified turnouts to yield a track-selective location system.

The remainder of this paper is organised as follows: Section 2 describes the railway infrastructure and basic structures of digital maps. The data models of the topological and geometrical level are presented in Sec. 3. The resulting digital map for a secondary line is shown in Sec. 4 while Sec. 5 concludes the paper.



2 Railway infrastructure representation

The infrastructure of a railway network is presented in different visualisation forms depending on the application, such as topological maps of the rail network for the dispatcher or geometrical maps for maintenance services. The following section describes different representation forms.

2.1 Topological representation of the railway network

The topological train station and rail network map contains not only the topological connections but it contains also further information such as signalling, turnouts, passenger platforms and level crossings. It is mainly used for dispatching purposes and by other staff members on the train station.

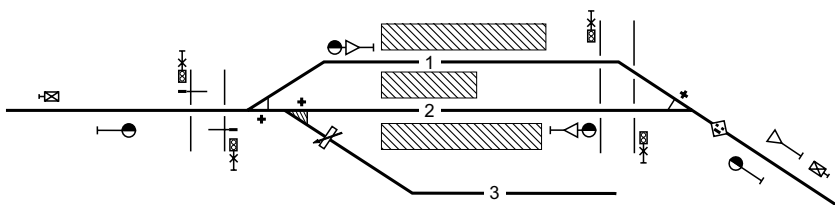


Figure 1: Topological example map of a train station and rail network.

2.2 Geometrical representation of the railway network

The geometrical representation of the railway network on a site plan is important for operational administration to organise maintenance and attendance services. On a site plan, all elements are represented geometrically correct in a defined coordinate system, e.g. Universal Transverse Mercator coordinate system. Such plans are edited with CAD programs or GIS systems and hence these data are already digitally available for use in a digital map.

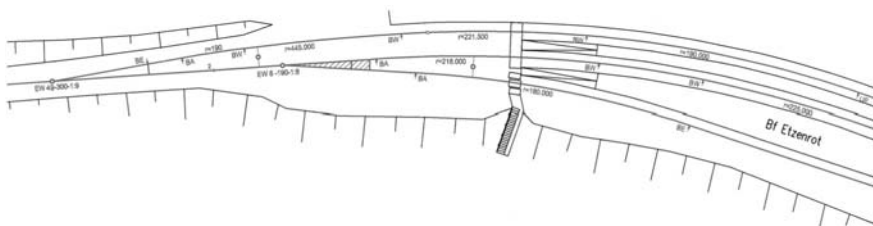


Figure 2: Example map of a geometrical train station site plan.



2.3 Basic structures of digital maps

The previous section describes different visualisation forms of the railway infrastructure with the goal to represent the infrastructure as detailed as necessary. In the following section, however, different models of digital maps are introduced. Also, various abstraction levels of the infrastructure are illustrated for different sensors and location methods.

One-dimensional maps are the most simple representation of the infrastructure with only one degree of freedom as displayed in Fig. 3. Thus a location system can be achieved by integrating the velocity and re-calibrating the system on discrete events, e.g. balises or turnouts.

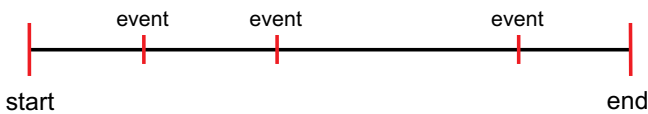


Figure 3: Linear map representing the railway network.

In topological maps the infrastructure is modelled by a node-link model, where each link is a one dimensional track element without a branching point. Changing the route is only possible on nodes as shown in Fig. 4. A location system can also function by integrating over train velocity. However, attention must be paid to path changes at nodes.

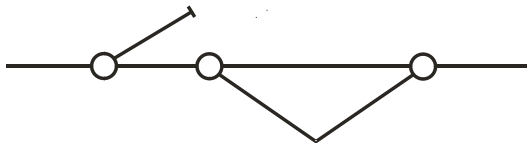


Figure 4: Topological map representing the railway network.

Basis for geometrical maps is also the node-link model, but in contrast to the topological map, all elements are displayed with their correct geometry and position. The geometrical track can be reconstructed by attaching all elements to each other. Train location is possible with all methods based on geometrical information.

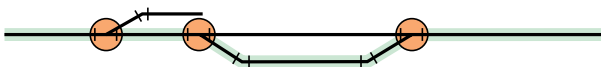


Figure 5: Geometrical map representing the railway network.

3 Data models of the railway infrastructure

Digital data sets can be modelled in different ways; often set theory is used. This mathematical formalism is however difficult to understand and are often less demonstrative. To overcome this drawback more simple formalisms like the Entity-Relationship Model (ER Model) can be used. With this model, the system architecture and data flow can be easily illustrated. The extension to the ER Model, the Extended Entity-Relationship Model (EER Model) has fixed semantics and a high expressiveness. Furthermore, it offers some constructs which do not expand the power of the model but increase its readability and manageability.

3.1 Extended Entity-Relationship Model

The EER Model is an extension to the ER-Model that was introduced by Chen [2]. Detailed definitions and descriptions of the EER Model explained on the basis of set theory are given in [3] and [4]. All fundamental terms necessary for modelling the digital map are explained in Fig. 6. Interested readers are referred to [5] and [6] for more information.

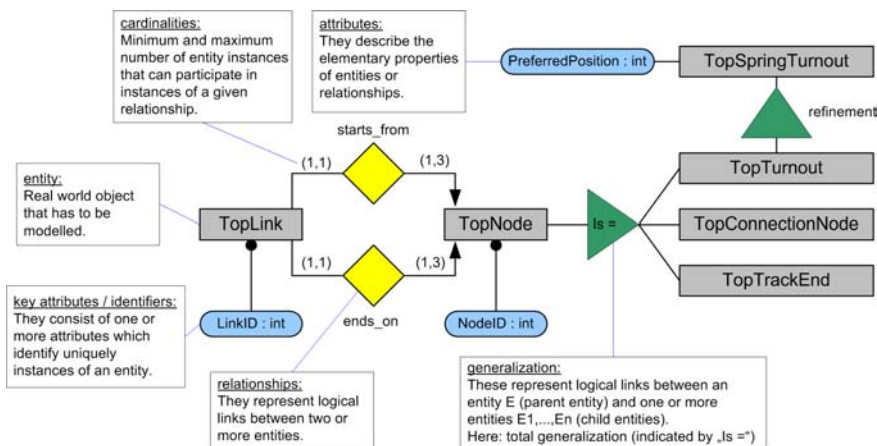


Figure 6: Introduction to the most important terms of the EER Model.

3.2 Modelling the topological level of the railway infrastructure

The topological level of the digital map reflects the railway infrastructure. Due to its net-like aspect, the concepts of graph theory can be used and thus nodes, links and the relationships between them must be defined. Figure 7 displays the node-link model of a single turnout with the associated right and left track and their continuation as suggested in [7] and [1]. The node *TopTurnout* represents a turnout as a zero-dimensional object that divides the rail track into two tracks.



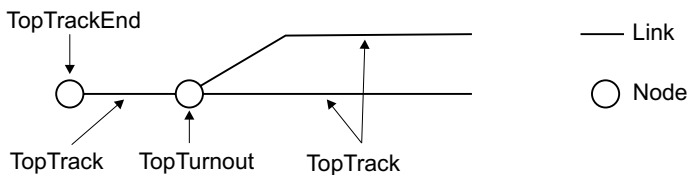


Figure 7: Representation of the topological level through the node-link model.

For a train location system that uses sensors which detect and classify turnouts [8], the node-link model of the topological level must be extended, so that turnouts are modelled as spatially expanded objects. One modelling approach is to divide the turnout into three elements: a branching point as node and both tracks as spatially expanded links. With this turnout representation, a sensor system, e.g. eddy current sensor system, can detect turnout elements, such as points, frogs or guard rails and can compare them with the turnout elements of the topological route map. An example for this extension displaying a single turnout is presented in Fig. 8.

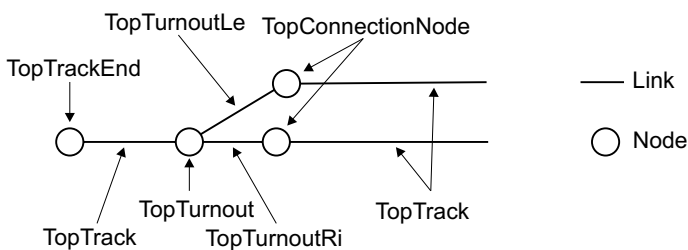


Figure 8: Extension to the topological node-link model for spatial modelling of turnouts.

3.2.1 EER Model of the topological level

By using the previously presented extension of the node-link model, an EER Model of the topological level can be designed as illustrated in Fig. 9.

3.2.2 Entity and relationship definitions of the topological EER Model

This section describes in detail the entities and relationships used in the topological EER Model shown in Fig. 9.

TopLink

The region of the digital map will be divided into an finite number of disjunctive elementary tracks. Each track is represented by the entity *TopLink* with the key attribute *LinkID*. The connection to the entity *TopNode* is modelled through the relationships *start_from* and *ends_on*.



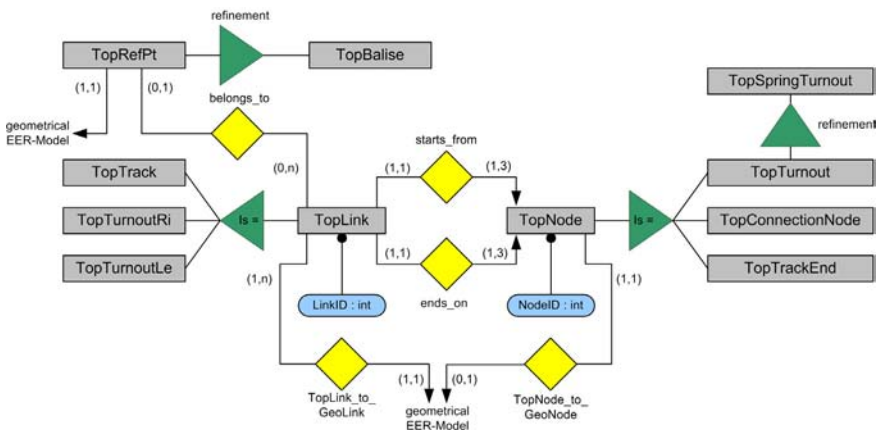


Figure 9: EER Model of the topological level.

TopTrack, TopTurnoutRi and TopTurnoutLe

The entities *TopTrack*, *TopTurnoutRi* and *TopTurnoutLe* are total generalisations of the entity *TopLink*. *TopTrack* represents the topological rail track. The entities *TopTurnoutRi* and *TopTurnoutLe* model the spatial expansion of the turnouts with the orientation, right or left, in the entity name.

The relationship *start_from* and *ends_on* between *TopLink* and *TopNode*

This relationship represents the connection between *TopLink* and *TopNode*, where each *TopLink* can be connected at minimum with one and maximum with three *TopNode* instances.

TopNode

Each *TopNode* is the juncture between one, two or three *TopLink* instances and has the key attribute *NodeID*.

TopTurnout, TopConnectionNode and TopTrackEnd

The entities *TopTurnout*, *TopConnectionNode* and *TopTrackEnd* are total generalisations of the entity *TopNode*. *TopTurnout* is connected with three, *TopConnectionNode* with two and *TopTrackEnd* with only one *TopLink* instance.

TopSpringTurnout

The entity *TopSpringTurnout* models a turnout with a predefined orientation clamped by a spring.

TopRefPt

This entity describes a reference point placed on a *TopLink* instance.

The relationship *belongs_to* between *TopLink* and *TopRefPt*

Many instances *TopRefPt* can be connected with one instance *TopLink*, but only one instance *TopLink* can be connected with each instance *TopRefPt*.



TopBalise

The entity *TopBalise* is an refinement of *TopRefPt* with additional attributes.

The relationships *TopLink_to_GeoLink* and *TopNode_to_GeoNode*

It describes the connection between the topological level and the geometrical level.

3.2.3 Consistency checks of the topological EER Model

For data verification and validation of the topological EER Model, several consistency checks must be implemented. The following section describes only a few checks to illustrate their use:

1. One instance of *TopConnectionNode* can connect min/max two instances of *TopLink*.
2. The sum over the *TopLink* instances that are connected with the relationship *start_from* and *ends_on* with a *TopNode* instance, is at minimum one and at maximum three.
3. Each instance of one entity must have a different key attribute.

And so on.

3.3 Modelling the geometrical level of the railway infrastructure

In the geometrical level, each elementary track of the railway infrastructure can be described by its geometrical position defined by start- and endpoint as well as its geometry. In this model, only three different geometrical types are distinguished: straight track, curved track and transition curved track. With the aid of the node-link model the geometrical representation of the railway infrastructure can be completely reconstructed.

Figure 10 illustrates the node-link model of a single turnout. If a link belongs to a turnout, an additional differentiation, e.g. “geometrical track turnout right” or “geometrical curved track turnout left”, is done to indicate the left or right route. A detailed description of the used entities is given in Sec. 3.3.2.

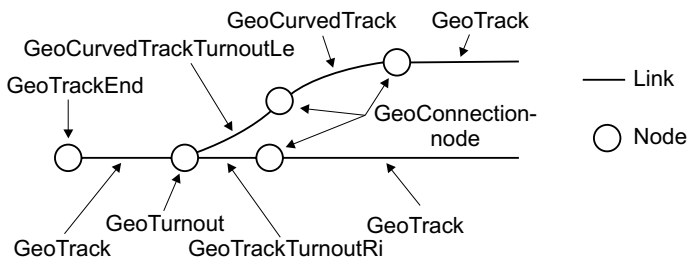


Figure 10: Representation of the geometrical level through the node-link model.



3.3.1 EER Model of the geometrical level

Applying the node-link model presented in Fig. 10 to the railway infrastructure, an EER Model for the geometrical level can be designed. For better readability, the relationships *GeoLink_to_TopLink* and *GeoNode_to_TopNode* connecting the geometrical and topological level are plotted in both EER Models.

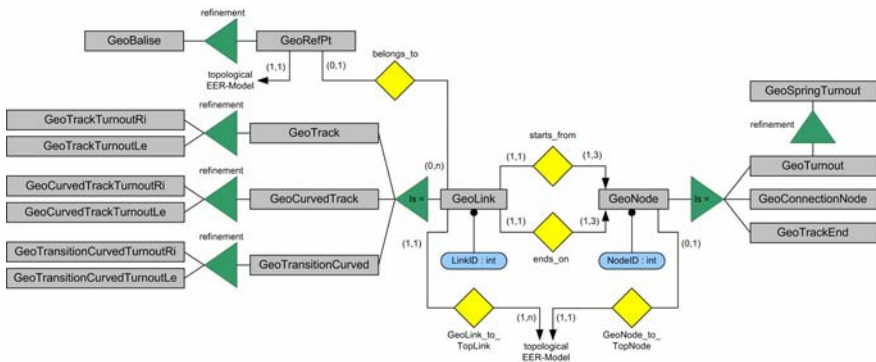


Figure 11: EER Model of the geometrical level.

3.3.2 Entity and relationship definitions of the geometrical EER Model

Due to the fact that both EER Models are constructed from similar node-link models only entities having major differences are explained in this section. Entities having minor differences are *GeoLink*, *GeoNode*, *GeoTurnout*, *GeoConnectionNode*, *GeoTrackEnd*, *GeoSpringTurnout*, *GeoRefPt* and *GeoBalise*. All used relationships having also only minor differences and thus their explanation is given in Sec. 3.2.2.

GeoTrack, *GeoCurvedTrack* and *GeoTransitionCurved*

These entities are total generalisations of the entity *GeoLink*. Additionally, each entity has further attributes describing its geometry such as radius, orientation and so on.

The refinement of the previously given entities into *GeoTrackTurnoutRi*, *GeoTrackTurnoutLe*, *GeoCurvedTrackTurnoutRi*, *GeoCurvedTrackTurnoutLe*, *GeoTransitionCurvedTurnoutRi* and *GeoTransitionCurvedTurnoutLe* are necessary to complement the possible routes over a turnout.

3.3.3 Consistency checks of the geometrical EER Model

Data verification and validation of the geometrical EER Model must be done in a similar manner as with the topological EER Model. Again, only a few checks are written down to illustrate their use:

1. Each start- and endpoint of a *GeoLink* must be on the position of a *GeoNode*
2. Each instance *GeoLink* must have a connection to an instance *TopLink*.

And so on.



3.4 Summarising the modelling of the railway infrastructure

With the introduced topological and geometrical models, a complete and consistent representation of the railway infrastructure is possible. The extension of the topological node-link model, which includes turnout components, is necessary for a train-borne location system that uses turnouts as reference points [9]. Furthermore, the combination of both levels enables the use of various location methods, where the topological and geometrical information are used to determine the train position.

4 Results

The following section describes the implementation of the digital map for the Albtalbahnhof (secondary line) in Germany. The input data to the digital map are gained by a surveying team and thus accuracy of only a few inches is achieved. With the use of a CAD/GIS program, the elementary tracks have been extracted and all necessary attributes are attached. After applying the consistency checks to verify the imported data, the topological and geometrical objects are exported into a MySQL data base. By using the EER Models, these objects can be easily mapped into the data base. A further advantage of using MySQL is the standardised interface into other software projects. Having all information of the railway infrastructure in the data base, the topological or both, the topological and geometrical levels can be used in location systems.

For the secondary line with a line mileage of 26 km, the implementation of the digital map results in about 1800 instances with 17500 attributes subdivided into 877 nodes and 911 links. Along this secondary line there are nine train stations and about 60 turnouts. Figure 12 shows a small cut-out of the digital map printed on a topographical map. The creation of the digital map for this secondary line by applying the presented models and exporting the objects to the data base indicates the successful modelling and implementation. Using this digital map in combination with an eddy current sensor system, a train location system for secure applications is obtained.

5 Conclusions

This contribution has proposed models for generating a digital map for secure applications. The map is subdivided into a topological and geometrical level to accomplish the use of both parts individual or together. With the use of the EER Model, the railway infrastructure and its system architecture can be modelled with fixed semantics. By extending the node-link model to represent turnouts as spatially expanded objects, it is also possible to use an eddy current system for detecting turnouts as reference points and hence achieve a location system for secure applications.



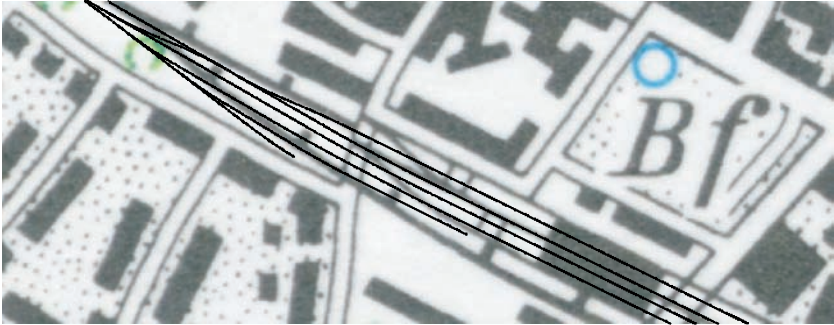


Figure 12: Visualisation of the geometrical level on a topographical map.

References

- [1] Bikker, G., Klinge, K.A., Schroeder, M. & Schnieder, E., Concepts of an intelligent route atlas for localisation in guided traffic. *4th International Conference on Vehicle Navigation*, EURNAV 98: Hannover, 1998.
- [2] Chen, P., The entity relationship model - towards a unified view of data. *ACM Transactions on Database Systems*, volume 1, pp. 9–36, 1976.
- [3] Hohenstein, U., *Formale Semantik eines erweiterten Entity-Relationship-Modells*. Teubner: Stuttgart/Leipzig, 1993.
- [4] Gogolla, M., *An Extended Entity-Relationship Model*. Springer: Berlin, 1998.
- [5] Maschek, U., *Datenmodell zur Planung von Stellwerken*. Ph.D. thesis, Technische Universität Braunschweig, Institut für Eisenbahnwesen und Verkehrssicherung, 2002.
- [6] Böhringer, F., Generierung einer digitalen streckenkarte zur ortung im schienengebundenen verkehr. Technical Report AV 02/05, Universität Karlsruhe (TH), Institut für Mess- und Regelungstechnik, 2005.
- [7] Gielsdorf, F., *Datenmodellierung für Gleisnetze und Schaffung eines einheitlichen Raumbezugssystems*, volume Reihe C, Nr. 491. Verlag der Bayerischen Akademie der Wissenschaften: München, 1998.
- [8] Geistler, A. & Böhringer, F., Detection and classification of turnouts using eddy current sensors. *Computers in Railways IX*, eds. J. Allan, R.J. Hill, C.A. Brebbia, G. Sciutto & S. Sone, WIT Press: Southampton, pp. 467–476, 2004.
- [9] Geistler, A., Train location with eddy current sensors. *Computers in Railways VIII*, eds. J. Allan, R.J. Hill, C.A. Brebbia, G. Sciutto & S. Sone, WIT Press: Southampton, pp. 1053–1062, 2002.



Train position detecting system using radio millimeter-waves

T. Maeda, K. Watanabe & M. Ono
Central Japan Railway Company, Japan

Abstract

The Maglev on the Yamanashi Test Line is driven by LSM (Linear Synchronous Motor). Train control needs a precise position of the train in order to achieve stable and efficient operation of the LSM drive. Moreover, the information of the train position is important from the viewpoint of safety, because it is used to check a train speed for safety brake control that assures the last method. In the Yamanashi Maglev Test Line, a train position is detected by the inductive radio system. The inductive radio system is reliable and accurate. On the other hand, the cost of construction and maintenance of the inductive radio system is expensive, because the cross-inductive cable has to be laid along the guideway accurately and the Local Position Detectors have to be installed about every 2km. Thus, the train position detecting system has been developed without inductive wire in order to reduce construction and maintenance cost. It detects the train position by a combination of several methods, such as an EMF (Electromotive Force) observer, a tachometer of the wheel, a counter of ground coils for levitation and guidance, and radio millimeter-waves. In this paper, out of those detecting methods, the train position detecting system using radio millimeter-waves is presented.

Keywords: radio system, millimeter-wave, position detecting system, maglev.

1 Introduction

Since 1997, running tests have progressed smoothly, for over 400,000km and setting a new world speed record of 581km/h, in Yamanashi Maglev Test Line. Recently, long-term endurance test has been performed and research and development activities have been continued in order to reduce the cost of the commercial system.



In the Yamanashi Maglev Test Line, a communication system between the Test Center and the Maglev vehicles uses an LCX (leaky coaxial cable) radio. But it is considered that the future train radio system will require a larger transmission capacity. So a radio system using millimeter-waves has been installed for experiment.

On the other hand the train position detecting system without the inductive wire has been developed in order to reduce construction and maintenance cost. It detects the train position by combination of several methods, such as an EMF (Electromotive Force) observer, a tachometer of the wheel, a counter of ground coils for levitation and guidance, and radio millimeter-waves.

One of these, the detecting system using radio millimeter-waves is difficult to achieve the same accuracy as that with the inductive wire. Nevertheless, it is able to detect the absolute position by radio propagation between the base and the mobile stations, thus it is considered that the detected position data can be utilized for block control. And it is inexpensive to introduce the position detecting system using radio millimeter-waves because an exclusive radio system is not necessary.

2 System outline

2.1 Principle of detection

The data between the mobile and the base stations is transmitted with multiframes which are sets of consecutive frames. Figure 1 shows the structure of the multiframe.

The mobile station detects a time lag between transmitter (TX) and receiver (RX) multiframe, and the information is transmitted to base stations at the particular channel time slots in the multiplex line. The base stations detect a time lag between TX and RX multiframe, and compare the time lag at the mobile station with that at the base stations, and calculate the distance between the mobile and the base stations. Then the train position is detected with the distance and the base station location. This method has been devised for being applied to independent synchronization system between the mobile and the base stations, such as the radio system using millimeter-waves in the Yamanashi Maglev test line.

Specifically, the train position is calculated by

$$P_m = P_b + D$$

$$D = T3 \times c = \frac{T2 - T1}{2} \times c \quad \text{when } T2 > T1$$

$$D = T3 \times c = \frac{T2 + Tm - T1}{2} \times c \quad \text{when } T2 \leq T1$$

$$(2 \times T3 < Tm)$$



where
 P_m : Position of Mobile Station
 P_b : Position of Base Station
 D : Distance from Base to Mobile Station
 $T1$: Time lag between TX and RX multiframe at Mobile Station
 $T2$: Time lag between TX and RX multiframe at Base Station
 $T3$: Radio Propagation time
 Tm : Time length of one multiframe
 c : Light speed

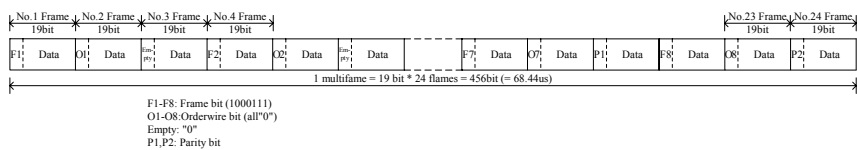


Figure 1: Structure of multiframe.

Figure 2 shows a chart of each multiframe. Figure 3 shows the position detecting flow.

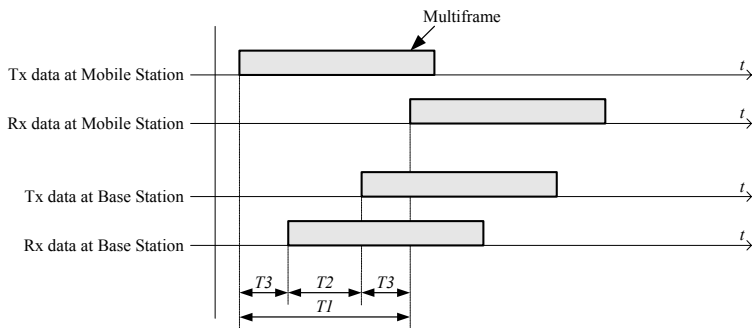


Figure 2: Chart of each multiframe ($T2>T$).

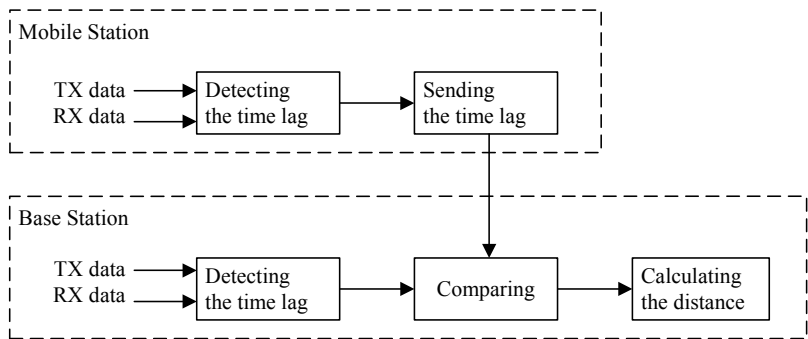


Figure 3: Position detecting flow.



2.2 Composition of position detector with radio system

Figure 4 shows the composition of the experimental system. The position detector obtains RX multiframe from DEM (Demodulate unit) and TX multiframe from TXBB (Transmitter baseband unit) in the radio device. The position detector is connected with the radio device for the time lag data and voice transmission at the speed of 6.312Mbps. At the base station, the position detector transmits the detected distance data to the Evaluation device. The Evaluation device obtains the standard position data from the position detector using the inductive radio system.

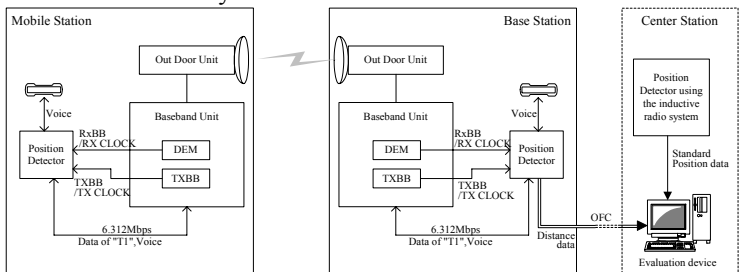


Figure 4: Composition of position detector with radio device.

2.3 Test system configuration

Figure 5 shows the test system configuration. In the Yamanashi Maglev Test Line, redundancy is considered in the design of the radio system using millimeter-waves. For example, any section of the test line is covered with three base stations. The base stations which communicate with the mobile station change as Maglev vehicles run. Thus, a communication between Maglev vehicles and the Test Center is kept continuously all along the test line.

In the present test line configuration, three base stations have been installed, each of which is equipped with one position detector. Thus, three kinds of position data are obtained at the same time.

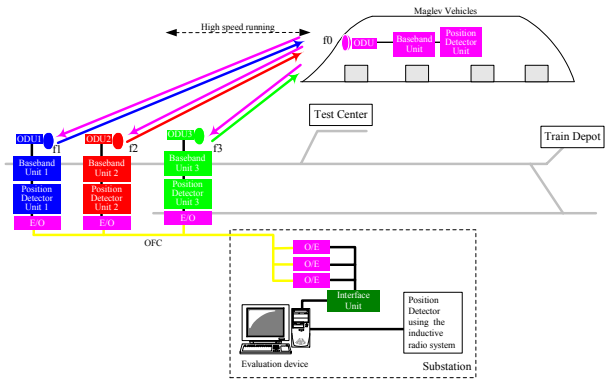


Figure 5: Test system configuration.



3 Test results

3.1 Running test results

The position data have been acquired at high speed running tests. Figure 6 shows an example of the results. From this, it has been confirmed that it is able to detect the train position continuously from three base stations up to about 10km that is the detecting limit at high speed.

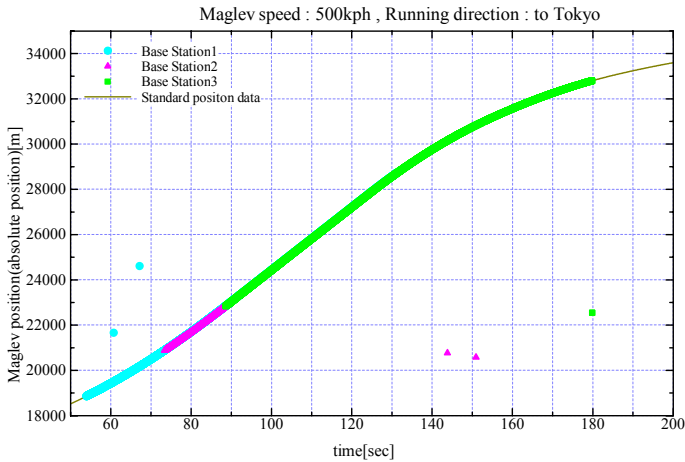


Figure 6: Example of detected position data by three base stations.

3.2 Analyses

3.2.1 Rate of ineffective data

The detected data contains extraordinary items because of transmission error or detecting fault. Here, the datum that deviates from the standard one by more than 10m is defined as ineffective, and the rate of ineffective data is calculated. As a result, the rate is as low as about 2%, thus it is considered that the rate level is not a problem as the position detecting tool. In addition, it has also been revealed that the rate of ineffective data does not depend on the train speed.

3.2.2 Characteristics of position error

It has been known that the position error increases as the mobile station goes away from the base stations. Thus, the characteristics of the position error from the standard data have been analyzed under every topography condition in the Yamanashi Maglev Test Line. The topography of the Yamanashi Maglev Test Line is described in Figure 7. Section classification in Figure 7 is explained in Table 1. The result of analysis is shown Figure 8. Consequently, a trend of the increase in position error is characterized for every topography condition. It is considered that the position data can be corrected based on the trend.



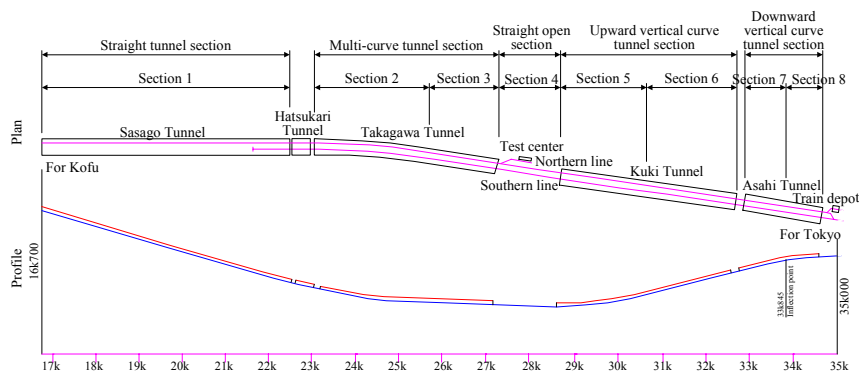


Figure 7: Topography of Yamanashi Maglev test line.

Table 1: Section classification.

Section No.	Classification
1	Straight tunnel section
2	Multi-curve tunnel section
3	Straight tunnel section
4	Straight open section
5	Upward vertical curve tunnel section
6	Straight tunnel section
7	Downward vertical curve tunnel section to inflection point
8	Downward vertical curve tunnel section from inflection point

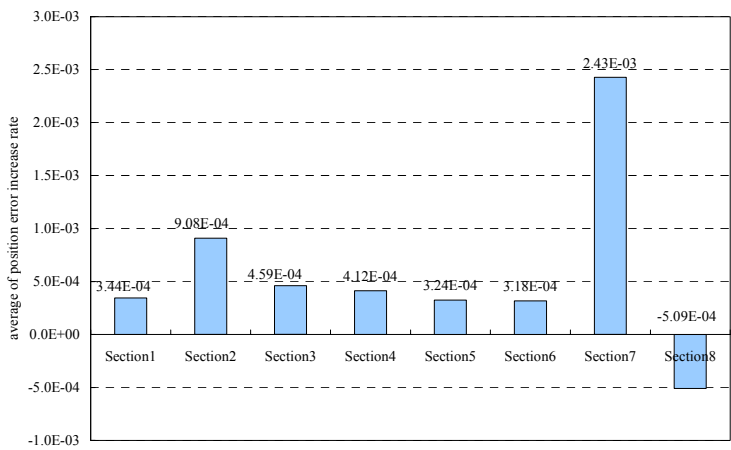


Figure 8: Characteristics of position error.



3.2.3 Results of processing data

The position data have been processed by the following procedures. The example of position error data before processing is shown Figure 9, and the example of position data after processing is shown Figure 10. The extraordinary data are interpolated, and the scatter of the position error data became considerably small.

(1) Position data smoothing

The position data are smoothed by the method of least squares.

$$y = a + bx$$

$$a = \frac{(\sum y_i)(\sum (x_i^2)) - (\sum x_i)(\sum x_i y_i)}{k(\sum (x_i^2)) - (\sum x_i)^2}$$

$$b = \frac{k(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{k(\sum (x_i^2)) - (\sum x_i)^2}$$

k : population parameter = 30

(2) Delay correcting

The position data are corrected using the delay time from the standard data.

(3) Position error correcting

The position data are corrected, using correction the data shown in Figure 8.

(4) Position data selection

The most suitable position datum is selected from three position data.

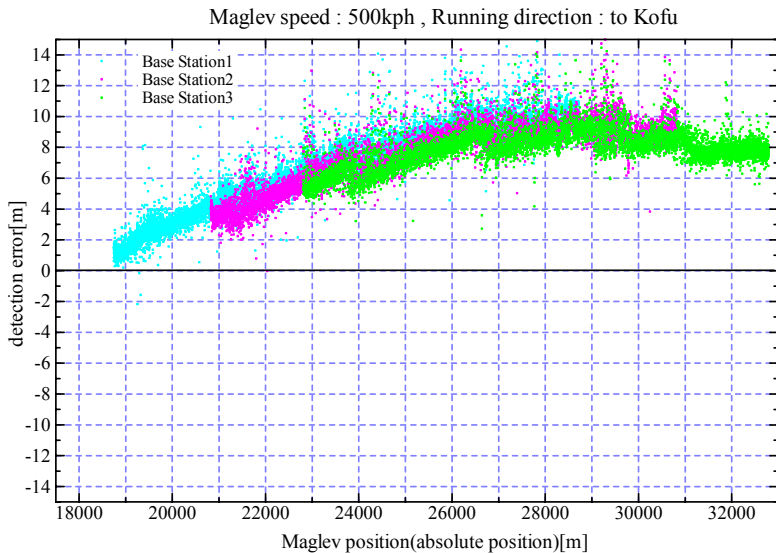


Figure 9: Example of position error before processing.



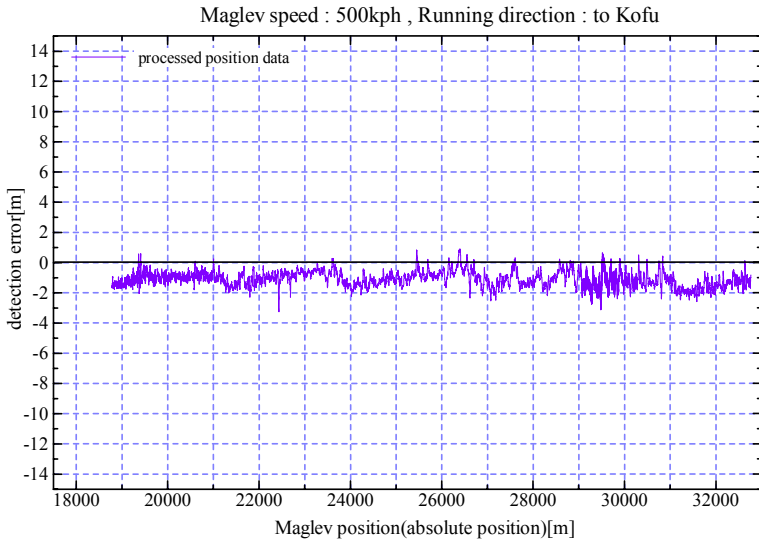


Figure 10: Example of position error after processing.

4 Conclusion

In the development, the knowledge of the characteristics of the position data has been obtained, and the method of processing of the position data has been established. It is considered that the system can be utilized as one of the position detecting tools.

The position detecting system will be installed along the whole Yamanashi Maglev Test Line, and will be developed aiming to realize practical use of a new position detecting system without the inductive wire.

References

- [1] H. Yamamura & S. Sasaki, Millimeter Wave Propagation Model and Delay Spread along the Maglev Guideway, *IEICE TRANS COMMUN*, vol.E78-B, No.8, 1995
- [2] H. Niikura, H. Yamamura, K. Yamakawa, H. Uenae & K. Honma, Train Radio System Using Millimeter-Waves for the Superconducting Maglev System, *Maglev '95*, PP405-409, 1995
- [3] T. Nakashima, S. Kaga & H. Ohtsuki, Results of electrical experiments on the Yamanashi Maglev Test Line, *Maglev '2000*, PP89-94, 2000
- [4] S. Koga, J. Enomoto & S. Kaga, Sensorless Control System of Superconducting Maglev, *Maglev '2002*, 2002
- [5] S. Miyamoto, Y. Osada, K. Yamazumi & T. Furuki, The Status of the running tests of JR-Maglev, *Maglev '2004*, PP60-64, 2004



Section 9

Dynamic train regulation

This page intentionally left blank

Optimal train control at a junction in the main line rail network using a new object-oriented signalling system model

R. Takagi¹, P. F. Weston², C. J. Goodman², C. Bouch³,
J. Armstrong⁴, J. Preston⁴ & S. Sone¹

¹*Department of Electrical Systems Engineering, Kogakuin University,
Japan*

²*Rail Research UK, School of Engineering
University of Birmingham, UK*

³*Rail Research UK, School of Engineering (Civil Engineering),
University of Birmingham, UK*

⁴*Rail Research UK, Transportation Research Group,
School of Civil Engineering and the Environment,
University of Southampton, UK*

Abstract

On a main line railway network with many junctions, the delay of a train is likely to cause delays to many other trains, especially because of conflicts at junctions. Optimising one junction, however, may have an adverse effect on other parts of the rail network because of the mixed-traffic situation of most main line railways. To approach the complicated problem of optimal re-scheduling in response to the delay of a train, an efficient algorithm must be sought.

The authors have taken a junction as an example, and have performed numerical optimisation on a case when the services through this junction are disrupted. The objective criterion is the weighted sum of train times. The optimisation program uses the Object-Oriented Multi-Train Simulator (OOMTS) developed by Birmingham University, as an embedded simulator. In the optimisation routine, a Genetic Algorithm (GA) was used to optimise the order of route setting.

In this paper, the authors give details of a model junction, and a brief explanation of the OOMTS. The authors then explain how a GA can be applied to solve this problem, especially the chromosomal expression of the problem. The results of numerical optimisations for different weighting parameters are shown based on which the authors discuss the feasibility of the proposed method. *Keywords:* object-oriented, multi-train simulator, railway signalling, junction, train pathing, optimisation, Genetic Algorithm.



1 Introduction

On a main line railway network with many junctions, the delay of a train is likely to cause delays to many other trains, especially because of conflicts at junctions. Junctions without flyovers, as are commonly seen in the British rail network, tend to make matters worse. Also, most main line railway lines run mixed traffic with a combination of fast and slow trains, making the junction management even harder. Therefore, controlling a junction in response to a disruption by minimising the accumulated delay over a sequence of trains at that point may still have serious adverse effects elsewhere in the rail network. Thus, the problem of optimal re-scheduling in response to a delay of a train is computationally a difficult task, and an efficient algorithm must be sought to approach this problem.

In an attempt to address this problem, the authors have taken a junction as an example, and have performed numerical optimisation on a case when the services through this junction are disrupted. The junction that has been used for the study is Abbotswood Junction on the Birmingham to Bristol Line in the National Rail network of the UK. The optimisation program uses the Object-Oriented Multi-Train Simulator (OOMTS) developed by Birmingham University [1], as an embedded simulator. In the optimisation routine, a Genetic Algorithm (GA) is used to optimise the order of route setting.

2 OOMTS: an outline

The first version of the Object-Oriented Multi-Train Simulator (OOMTS) at the University of Birmingham was written by Siu [2] in the early 1990s. The logic of the program is based on a previous Fortran version [3], under constant development since 1973 and used by many rail operators and manufacturers around the world.

The model implemented in OOMTS comprises of five “subsystem managers” [4], each containing data (or the instances of objects necessary for running the simulation) for a railway “subsystem”. The Simulation Manager is at the top of the structure, managing all these subsystem managers and the actual running of the simulation itself.

The following are brief explanations of the five subsystem managers:

1. Network Manager: This subsystem manager holds instances of objects containing data for the topography and track profile (gradient, curve radius and line speed) of the rail network to be simulated.
2. Signal Manager: This subsystem manager holds instances of objects containing data for the signalling system.
3. Rolling Stock Manager: This subsystem manager holds data for the types of rolling stock to be used in the simulation.
4. Power Network Manager: This subsystem manager holds instances of objects for the electric power supply system.
5. Train Service Regulator: This subsystem manager holds instances of objects containing data for various aspects of train services, including timetable information.



OOMTS was originally designed and tested for metro lines and suburban lines, for example the Island Line of Hong Kong's MTR Corporation, and has successfully been used by researchers at the University of Birmingham, latterly within Rail Research UK, the universities' centre for railway systems research [5].

3 The model used in the optimisation

Abbotswood Junction, on the Birmingham to Bristol Line in the UK, is used as the model on which the optimisation was performed.

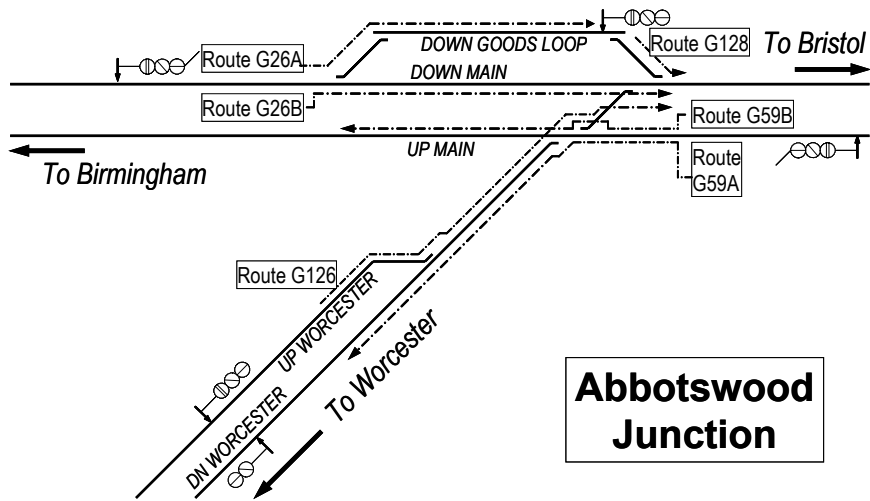


Figure 1: Abbotswood Junction with six possible routes.

Figure 1: shows the schematic of the junction. The Birmingham to Bristol Line is double track, and a single track branch line from Worcester joins here. Another feature of this junction is the Down Goods Loop, which allows passenger trains from Birmingham to Bristol Line to overtake a freight train. There are six possible routes for trains, and no reversing of trains is allowed.

The junction is modelled using the actual detailed data of the signalling system that is currently in use. This detailed data also covers the nearby part of the network, namely:

- 1) about five kilometres from Abbotswood Junction towards Birmingham,
- 2) about five kilometres from Abbotswood Junction towards Bristol, and
- 3) about 500 metres from Abbotswood Junction towards Worcester.

For the optimisation study, ten kilometres of double-track have been added to both the Birmingham and the Bristol end of the Birmingham to Bristol Line track data, and one kilometre of double-track has been added to the Worcester end of the branch line data. Virtual stations are defined at the far ends of each of these tracks, and trains are assumed to travel between these stations.



Figure 2: shows the assumed schedule of trains run through Abbotswood Junction. The trains involved are: one freight train on the Down Main track from Birmingham to Bristol; three passenger trains each for both Down and Up Main lines between Birmingham and Bristol; and three passenger trains each for both Down and Up Worcester tracks between Worcester and Bristol.

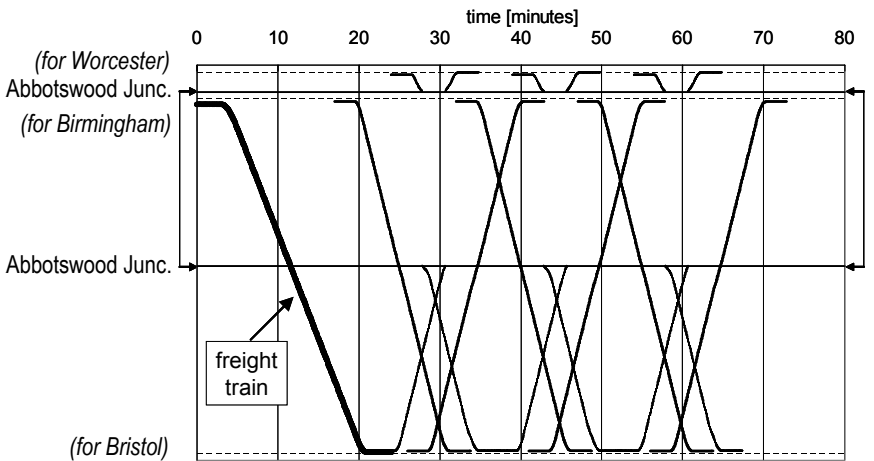


Figure 2: Assumed schedule of trains including one freight and twelve passenger trains.

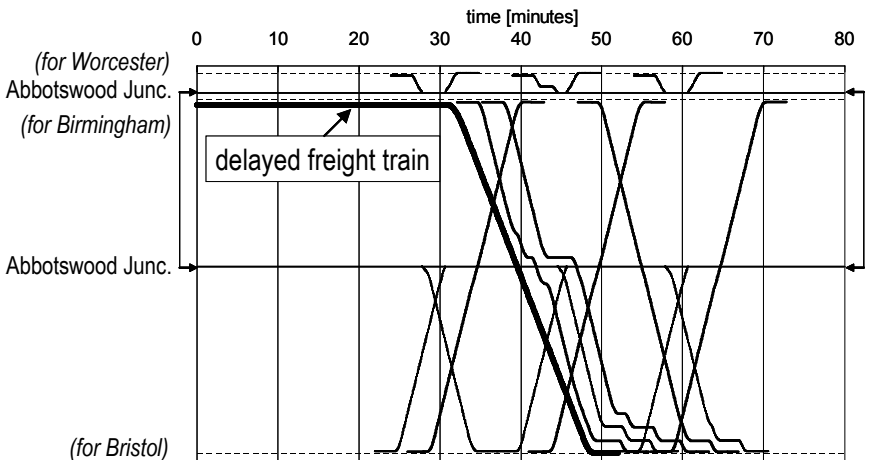


Figure 3: Result of applying “first-come, first-served” junction control with a 30-minute delayed freight train.



4 Assumed incident and the response by the non-optimal controller

The authors have assumed that the departure of the freight train is delayed for 30 minutes, causing the following Birmingham to Bristol passenger services to suffer reactionary delays.

Figure 3: shows the result of applying “first-come, first-served” type control, which is by no means optimal in this situation. Because the freight train is not scheduled to use the Down Goods Loop at Abbotswood Junction, the slow freight train causes more delay to the subsequent passenger trains at their destinations. There is also room for improvement by re-ordering passenger trains.

It is obvious that, provided the passenger trains have at least as high a priority as the freight train, the optimal solution must include the use of the Down Goods Loop by the freight train, which will enable one or more passenger trains to overtake the freight train. Then the decision must be made as to which of the following passenger trains should overtake the freight train.

5 Application of Genetic Algorithm

The authors have applied a Genetic Algorithm (GA) [6, 7] to identify the optimal control strategy for this situation.

To apply a GA to any optimal control problem, one has to express the control inputs as a “chromosome”, for which an evaluation can be given. For the most basic variation of GA, known as Simple GA (SGA), a chromosome must be a fixed-length binary sequence. Also, one has to define genetic operators; in the case of SGA they are mutations, crossovers and selections.

In the problem of train control at junctions, the control inputs are the setting of routes. The fitness of a particular routing sequence will be based on the time intervals between the scheduled departure times and the simulated arrival times, with weighting according to train priority. Here, it is assumed that the departure times and the performance of trains are given, which means that a set of control inputs can be defined as the order in which routes are set. One possible chromosomal expression, therefore, is the queue of routes; one example in the case of Figure 1: is (G26A, G26B, G128, G126, G59B ...).

However, using the queue of routes as the chromosomal expression has disadvantages. In the case of Figure 1:, if the head of the queue is G26A and the first train to arrive on the Down Main track from Birmingham is a passenger train (not allowed to go into the goods loop), then the system comes to a deadlock situation. Also, if there is already a train in the goods loop and the head of the queue is G26A, then the system comes to a deadlock situation because two or more trains are not allowed to be in the goods loop at the same time.

To avoid these disadvantages, the authors propose the use of a state-space trajectory as the chromosomal expression. In terms of control of railway junctions, the state of the rail network can be expressed as the queue of trains on any of the tracks that connect junctions. In the model used in this paper, the order



in which trains arrive at Abbotswood Junction is fixed, and the state expression can be simplified to a combination of four integers, as shown in Figure 4. Using this state expression, a chromosome can be defined as a sequence of system states, starting from the initial state, $(A, B, C, D) = (4, 3, 6, 0)$, and ending at the final state, $(A, B, C, D) = (0, 0, 0, 0)$. A change from one state to another indicates that a route is set. For example, in Figure 1:, the change of state from $(4, 3, 6, 0)$ to $(3, 3, 6, 1)$ means that the route G26A is set. Taking constraints into consideration is easier; for example, the constraint that two or more trains cannot be in the Down Goods Loop at any one time can be easily expressed as $D \leq 1$.

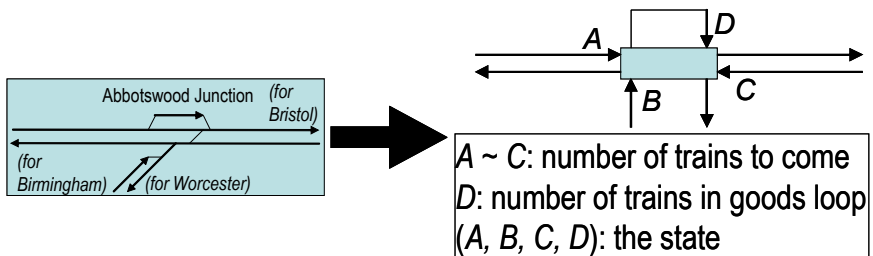


Figure 4: The state of the network.

A mutation operation can be defined as a partial change in the trajectory. Upon mutation the constraints can be taken into account so that the chromosome after the mutation is still a “feasible” solution. The probability of the occurrence of a mutation is to be higher when the mutation transforms an original chromosome to another one that has a smaller Levenshtein distance [8] to the original.

A crossover operation between two chromosomes is possible only when there are one or more common states (other than the initial and the final states) in the chromosomes. If there are two chromosomes P (sequence $S_{P,1}, S_{P,2}, \dots, S_{P,N}$) and Q (sequence $S_{Q,1}, S_{Q,2}, \dots, S_{Q,M}$), and if $S_{P,i} = S_{Q,j}$, then the crossover of these two chromosomes at $S_{P,i} = S_{Q,j}$ creates a new chromosome which has the sequence of either $(S_{P,1}, S_{P,2}, \dots, S_{P,i}, S_{Q,j+1}, S_{Q,j+2}, \dots, S_{Q,M})$ or $(S_{Q,1}, S_{Q,2}, \dots, S_{Q,j}, S_{P,i+1}, S_{P,i+2}, \dots, S_{P,N})$. If no common state is found in the sequences of P and Q except for the initial and final states, then the pair of chromosomes P and Q are called “infertile”.

Using this chromosomal expression, the problem can be solved in broadly the same way as the SGA.

6 Optimisation results and discussion

The authors have performed numerical optimisation on five different cases, using GA with the proposed chromosomal expression as explained in the previous section.



For all cases, a weighted sum of train times is used to evaluate a chromosome. The only difference between cases is the values of the weighting coefficients for passenger and freight trains. All other conditions are unchanged, remaining as described in the earlier sections.

Also, the parameters for the GA are common for all cases. There are 60 chromosomes per generation, and the algorithm terminates when 100 generations have been calculated. Given the population and evaluation of the current generation, the next generation of populations is determined as follows.

- 1) If the best evaluation among the current generation is worse than the very best which existed in the past, add the very best chromosome to the current generation population.
- 2) Create a group of 18 chromosomes by crossovers. One chromosome is created by the crossover between two parent chromosomes selected from the current generation, and is added to the new group. A chromosome in the current generation with better evaluation is assigned a higher possibility of being selected as one of the parents. If the selected pair is “infertile” and cannot create a crossover “child”, then mutation is applied to one of the pair and added to the new group instead.
- 3) Create the group of 42 chromosomes by selecting from the current generation of population, allowing the selection of the chromosome twice or more. The possibility that a chromosome in the current generation is selected is the same as the function used in 2).
- 4) Add the groups of chromosomes generated by 2) and 3) to create a group of 60 chromosomes.
- 5) Apply mutation to 80 per cent of the chromosomes in the group created in 4).

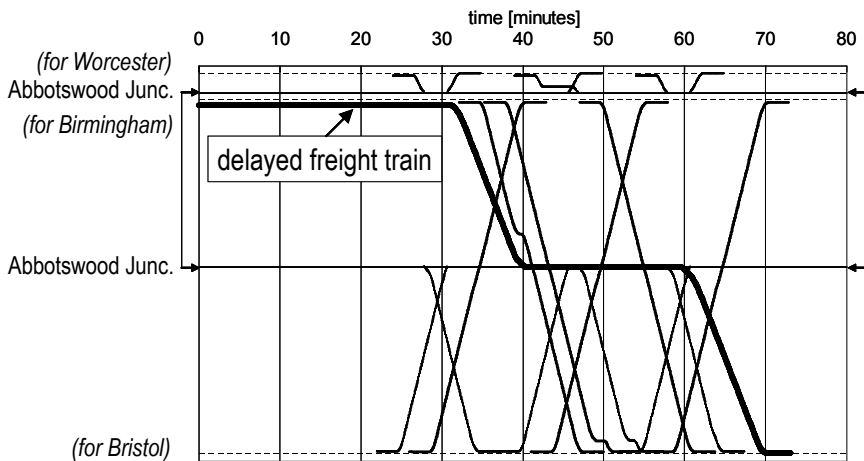


Figure 5: Optimisation result (1) weight of passenger : freight trains = 6 : 1.



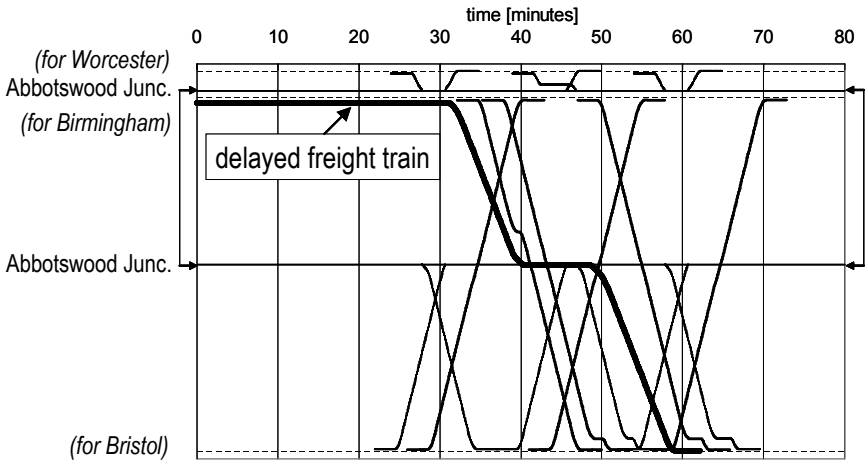


Figure 6: Optimisation result (2) weight of passenger : freight trains = 6 : 3 or 6 : 6.

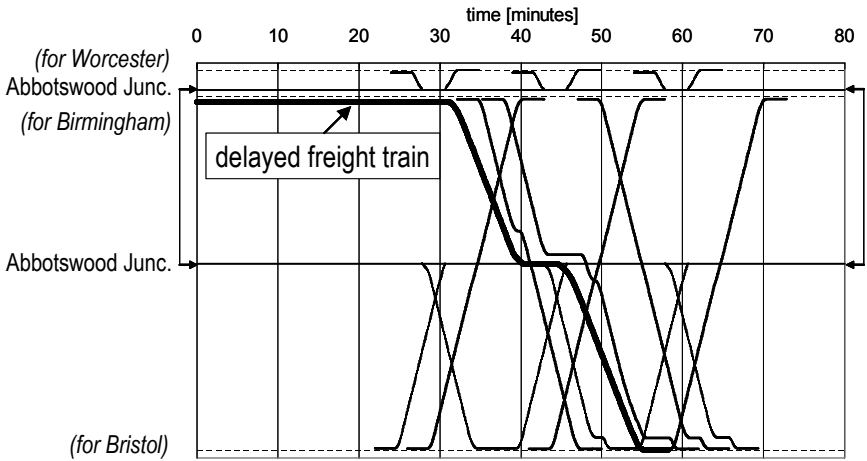


Figure 7: Optimisation result (3) weight of passenger : freight trains = 6 : 12.

Figure 5: shows the case where the weightings of passenger : freight trains is 6 : 1, in which the freight train departs the Down Goods Loop after all passenger trains bound for Bristol have made their way through Abbotswood Junction. Cases for the weightings of 6 : 3 and 6 : 6 yield the same result, shown in Figure 6:, in which the freight train leaves the last two Bristol-bound passenger trains behind. Results for the 6 : 12 and 6 : 24 weightings (shown in Figure 7: and Figure 8:, respectively) show that, as the weight for the freight train increases, fewer passenger trains overtake the freight train at Abbotswood, with Figure 8: showing that no overtaking becomes the optimal solution. These results suggest



that a change in the order of passenger trains and the freight train takes place when, as the weightings given for the freight train becomes larger, the disbenefit of holding the freight train is larger than the accumulated benefit of letting the passenger trains overtake at Abbotswood Junction.

The calculation time required for the optimisation was nearly 5 hours using a personal computer with 1.4GHz Intel Pentium 4 CPU; although this is clearly too long to be practical, considering that a detailed dynamic multi-train simulator has been used embedded in the optimisation process, it does not necessarily mean the GA cannot be used for this kind of optimisation.

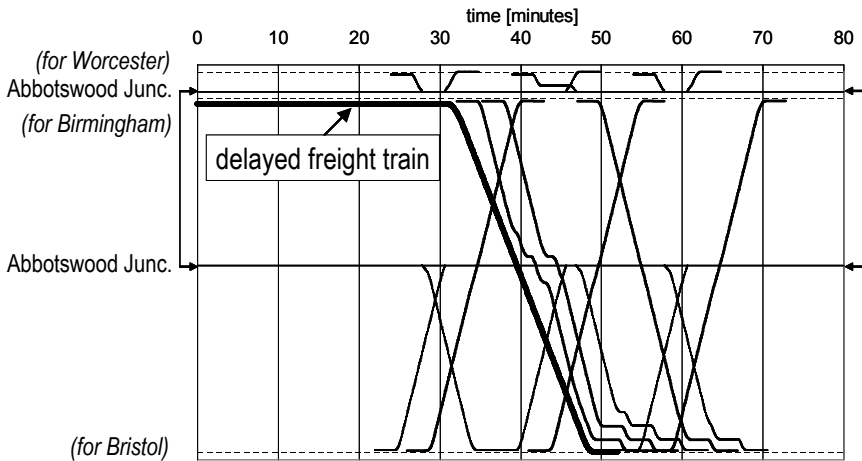


Figure 8: Optimisation result (4) weight of passenger : freight trains = 6 : 24.

The different results presented in this section are generated only by changing the weighting coefficients. However, there are many parameters that have to be determined, such as the size of the population per generation, percentage of chromosomes to be created through crossovers or percentage of chromosomes to mutate. These parameters may depend on track topography, the train schedule and the disruption scenario.

7 Conclusion

A number of numerical optimisations of a railway junction control have been carried out using Genetic Algorithms and using OOMTS as an embedded simulator. A novel chromosomal expression for the problem has been proposed that enables the problem to be solved in a way very similar to that used by the Simple GA. The authors are looking at the application of the same idea to various other problems, especially railway networks with multiple junctions.



Acknowledgement

This research was conducted within RRUK, the Universities' Centre for Railway Systems Research, funded by the EPSRC, U.K. (GR/S12784/01).

References

- [1] Takagi, R., *et al.*: "An Object-oriented Signalling System Model for the Study of Delay Minimization by On-line Train Control", *IRSE Aspect 2006*, London, UK (March 2006).
- [2] Siu, L. K.: "An Object-oriented Railway System and Power Network Simulator", The University of Birmingham (PhD Thesis) (1995).
- [3] Mellitt, B., Goodman, C. J. and Arthurton, R. I. M.: "Simulator for studying operational and power-supply conditions in rapid transit railways", *Proc. IEE*, Vol. 125, No. 4, pp.298-303 (1978).
- [4] Siu, L. K. and Goodman, C. J.: "An object-oriented concept for simulation", *Proc. COMPRAIL 1992 International Conference*, Washington DC, USA (1992).
- [5] Goodman, C. J.: "Promoting railway research in United Kingdom universities: the genesis of Rail Research UK", *Proc. of a Symposium on Research and Development in Railway Engineering*, Dept. of Elec. Eng., Hong Kong Polytechnic University, 3.1.1 - 3.1.12 (2003).
- [6] Nagao, T.: "Optimization Algorithms" (in Japanese), Shoko-do Publishing, Tokyo, Japan (2000), Chapters 10 – 12.
- [7] Rothlauf, F.: "Representations for Genetic and Evolutionary Algorithms", Physica-Verlag, Heidelberg, Germany (2002).
- [8] National Institute of Standards and Technology (NIST): "Dictionary of Algorithms and Data Structures" (<http://www.nist.gov/dads/HTML/Levenshtein.html>).



Optimising train priorities to support the regulation of train services with the assistance of active and deductive databases

C. Sakowitz & E. Wendler

Verkehrswissenschaftliches Institut, RWTH Aachen University, Germany

Abstract

Maximisation of revenues is a fundamental goal of any business-driven railway infrastructure company. In order to achieve this target in the context of traffic regulation, it must try to avoid delays and ensure scheduled connections. However, nominally equal delays to two different trains are not equal in value from an economic point of view in most cases. Moreover, some connections between trains might be more important in this sense than others. There are complex interdependencies and reciprocal effects in railway traffic. Considering these effects, a dispatcher must evaluate possible forms of conflict resolution and the waiting times these give rise to and select the best solution possible. This is not achievable where a time-critical conflict arises at short notice. Even closed mathematical optimisation algorithms encounter limits in the case of larger railway networks due to the enormous number of constraints to be considered. This paper will therefore propose that the optimisation process be separated from the train regulation process. Instead, economically evaluated train priorities for conflict situations are to be determined with the help of active, deductive and normative rules. Existing concepts of “smart” database management systems (DBMS) with integrated active and deductive database functionalities can be used for this application. An active DBMS allows the definition of reactions to be automatically initiated by the DBMS in response to the detection of given database-related events. A deductive DBMS allows new, deducible facts to be specified, administered and specially derived from explicitly introduced facts. Train priorities are generated and assigned in detachment from day-to-day operations for lightly and heavily disrupted railway traffic respectively. Long-term optimisation of these priorities is effected by evaluating past operational data.

Keywords: traffic regulation, real-time rescheduling, decision support system, train priorities, active databases, deductive databases.



1 Introduction

Given the possibility of sudden disruption or failure in railway operation, it is not always possible to run trains as foreseen in the timetable. The task for train dispatchers, therefore, is to promptly adopt suitable measures in the event of any disruption with a view to returning to normal service as soon as possible. The main objectives involved are twofold, namely to a) minimise delay minutes for trains and b) reduce passenger delay minutes over the entire journey chain. Failures in rolling stock, lineside equipment or the track itself lead to specific conflicts between trains which the dispatcher is required to resolve by adopting suitable measures. Priorities play a key role in how such measures are defined.

The conflicts concerned can be divided into two main categories, cf. [1]. Firstly, there are track-occupation conflicts between at least two trains. Owing to the guided motion principle underpinning the railway and to the system of technical protection employed, block sections with fixed boundaries can only be occupied by one train at most at any one time. Where several trains wish to have recourse to the same track-occupation entity simultaneously, a conflict arises that has to be resolved by prioritising one or other of the trains concerned. Lower-priority trains will then suffer an increase in running time as a result that is generally of the order of several minutes. Possible means of resolving the conflict at the expense of the lower-ranking train include switching it to an alternative route, extending a scheduled stop, relocating passing stops, adding an additional stop for operational requirements or increasing its running time, depending on what the situation calls for. A specific variant of these conflicts are track closure conflicts. This is a variation on the conflicting-route scenario in which a track-occupation entity cannot be requested for a specified period due to "external causes" (e.g. servicing work). The necessary response to such a conflict is as for a track-occupation conflict, with the train having by definition to be regarded as "lower in rank" than the external event. Secondly, at larger interchange stations in particular connectional conflicts can arise between trains. Connections serve the purpose of transferring passengers or wagons from a feeder train to one or more connecting services. If the feeder train is delayed by a few minutes, it may no longer be possible to effect this interchange without putting back the time of departure (and hence increasing the running time) of the connecting service. It is then necessary to accord priority either to the passengers or cargo on the feeder train (hold the connection) or to the punctuality of the connecting service (abandon the connection). If the decision goes against the connecting service, the latter's running time may be increased by several minutes. If the decision goes against the feeder train, passengers wishing to change trains may experience delays of up to an hour or more over the entire journey chain. Where freight operations are concerned, there may be delays of several hours in transfers of wagons at marshalling yards. A specific connectional conflict scenario takes the form of the circulation conflict. Here it is a case of the train connecting "with itself", as happens with a push/pull train at its terminus. What obviously distinguishes such connections is that breaking them is not feasible for technical reasons. One means of resolving conflicts may be to relocate the turnround to a previous station by omitting stops so as to allow



the connecting service to resume running under the original schedule. Here, too, a decision has to be made on whether to accord priority to the punctual running of the connecting service or to all scheduled stops being called at. Longer journeys for one of two carriage parties are at stake here again.

To be able to optimise priority-based conflict resolution procedures of this sort, a suitable rating criterion is required. Worthy of consideration is a performance regime of the type already adopted by one or two European railways and whose Europe-wide introduction is currently being discussed by the International Union of Railways (UIC). A regime of this kind can help quantify the weighting of punctuality and connectional certainty for a train in terms of transport economics. It is possible in this way to establish a punctuality rating for any scheduled stop by a train as a function of delay minutes. It is likewise possible to determine the penalties for connections not kept. Using the criterion of said performance regime as a basis, the decisions regarding priorities ought now to be specified in such a way that the resultant conflict resolution leads to ratings that are as low as possible.

Although there appears to be a case for resolving the problem adumbrated by adopting mathematical optimisation procedures and approaches of this kind are often proposed in literature in the field (closed linear optimisation, e.g. [2] and overview in [3], enumeration methods, e.g. [4]), their practical applicability is limited since, owing to the large number of constraints to be taken into account, real-time decisions can only be made in respect of comparatively lightly loaded networks with relatively straightforward topologies due to the exponentially rising computing input involved. Other approaches propose simplifying the constraints (e.g. doing away with difficult-to-manage constraints by occupying sets of points at interchange stations) so as to arrive at a solution that is theoretically optimal but cannot be put to effect on the actually existing network.

Technology is nevertheless sufficiently advanced in principle at present to allow conflicts to be automatically detected and resolved, fig. 1. Already today, adopting heuristics allows priorities to be used for the purpose of optimising conflict resolution within the context of simulation procedures [5]. These priorities are currently static and are prescribed manually. It is not possible to adequately render the fluctuating significance of a train movement in terms of transport economics by this means. Performance regimes have only been in place at railways for a short while (if at all) and thus are not yet made direct use of for the purpose of optimising the regulation of services.

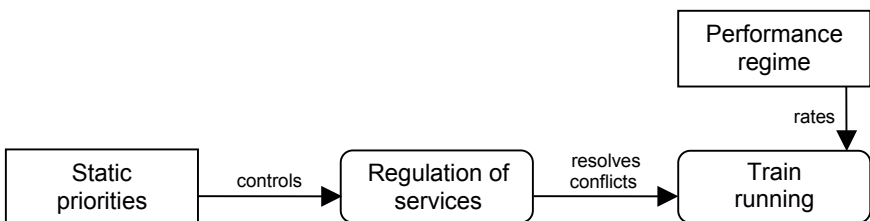


Figure 1: State of the art.



In the context of this paper, therefore, it is proposed reflecting advances made in technology by automatically adapting dynamic priorities in such a way that regulation can be optimised on the basis of a performance regime in future, fig. 2.

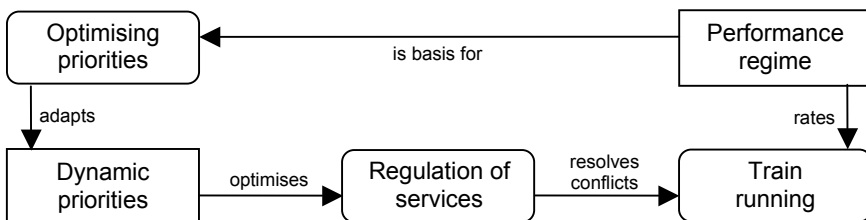


Figure 2: Proposal for new procedure.

Accordingly, a system for resolving conflicts is proposed in the following that combines the benefits of heuristic procedures (low computing time) with those of closed optimisations (quality achieved in rating traffic in terms of transport economics under the prescribed performance regime). To this end, the processes of priority-based traffic regulation, rating the outcome of regulation using the performance regime and the actual (longer-term) optimisation of priorities are divided into separate subsystems. Recourse can be had to “smart” database management systems for the optimisation process that merge deductive and active database techniques. Priorities are adapted in detachment from the day-to-day, time-critical conflict detection and resolution business of the dispatcher.

2 Modelling priorities

Priorities find application in the resolution of track-occupation, connectional and circulation conflicts by the regulation system, in which context it is also possible to regard circulation conflicts as being a special type of connectional conflict with fixed physical interlinking. It is also possible, by turn, to differentiate between train and connectional priorities.

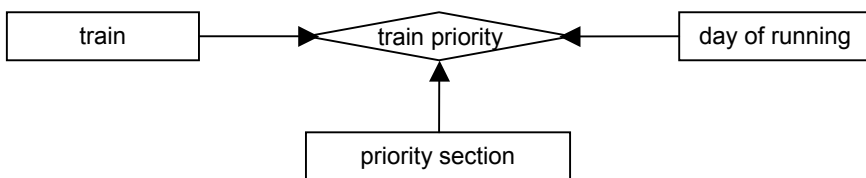


Figure 3: Train priorities.

Train priorities (see fig. 3) are defined for a train along specified, clearly demarcated sections of infrastructure on a given day of running, cf. [6]. Serving

as the boundary for such a “priority section” is a station having sufficient buffer capacity to facilitate the requisite waiting events given a change of train priorities. Each train is accorded a specified and unequivocal ranking on a given priority section from which operational precedence amongst trains involved in any track-occupation conflict in that priority section can be derived. The higher-ranking train involved in the conflict is given precedence, whilst the lower-ranking train is forced to wait in some way. Train priorities can be used to straightforwardly model lines on which certain types of traffic (long-distance passenger services, local passenger transport, freight) have precedence, a phenomenon that is steadily gaining in significance. Moreover, it is relatively easy in this way to take account of how the importance of a train movement fluctuates in transport economics terms in the course of its taking place.

Connectional priorities (see fig. 4) are defined for a specific connection between two trains at a station. They involve specifying maximum permissible waiting periods (beyond the scheduled time of departure) for the train maintaining the connection. It is possible by taking the delayed time of arrival of the connectional feeder train, the minimum interchange time and the maximum permissible waiting period to straightforwardly establish whether a connection is to be held or whether the connecting train is to depart on time - without waiting for the connection. The same applies by analogy for circulation transfer runs of stock from one train to another.

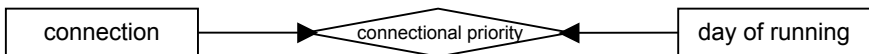


Figure 4: Connectional priorities.

3 Description of the overall architecture

As set out in the introduction, the overall architecture can be divided up into three separate subsystems existing alongside the outside world, specifically a regulation system, a monitoring system and a priority optimisation system (see fig. 5). The outside world represents actual train running on the existing infrastructure (railway operations). Elementary faults can arise on vehicles and the track that may result in conflicts that have to be remedied on a priority basis within the framework of traffic regulation. In addition, predefined types of traffic advice (notably automated or manual running advice) are generated and signalled in advance. The regulation system serves to detect and resolve actual conflicts on the basis of priorities prescribed by the priority optimisation system. The outcome of regulation is notified to the outside world for implementation. The regulation system also notifies the priority optimisation system of any advice of failure that may have a bearing on the according of priorities. Conflicts can in principle be resolved either manually, by centralised computer-aided means (train regulation centres) or by decentralised means (distributed regulation in large networks) using the priorities given. In the overall context, there is additionally a monitoring system for the purpose of assessing traffic against the



values prescribed under the performance regime. Standardised observations in this respect are likewise notified to the priority optimisation system so any adaptation of priorities required can be carried out. Priorities are generated by the priority optimisation system on the basis of input events conveyed and are supplied to the regulation system in good time. This process takes place with the help of rules.

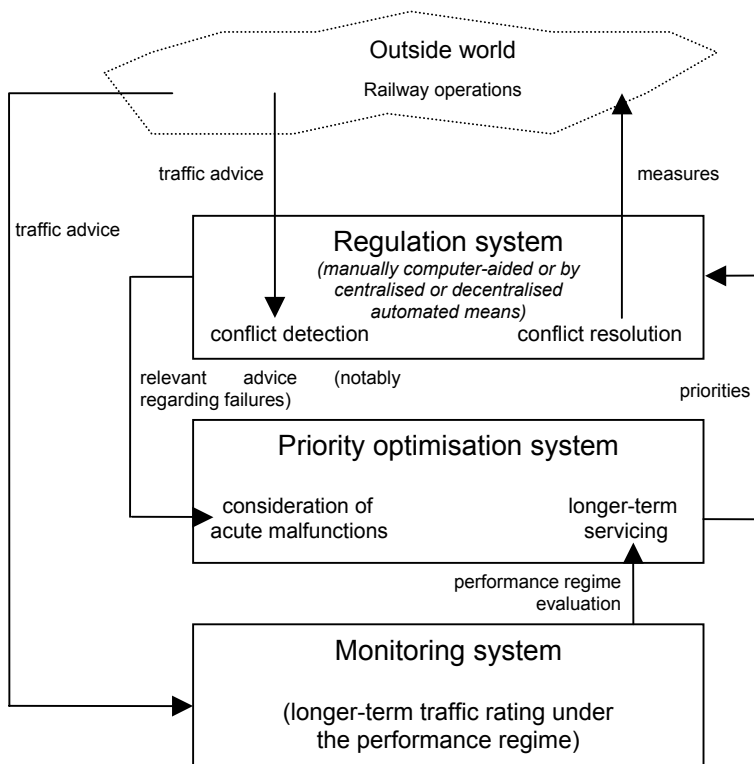


Figure 5: Overall system configuration.

The subsystems operate independently of one another and will be discussed in detail in the sections that follow. It is first intended to enumerate requirements for the regulation and monitoring system that are key to optimising priorities. Building upon this, important aspects of the priority optimisation system itself, which can be supported by both deductive and active database technology, will then be set out.

4 Key requirements for the regulation system

The (priority-based) regulation system has the function, in the event of traffic being regulated, of fine-tuning conflicting train paths by automated or manual



computer-aided means on the basis of the original schedule, the priorities prescribed and traffic advice from the outside world. Key constituents of the regulation system are an advice filter for incoming traffic advice, a running-time forecasting computer plus the functions of conflict detection and resolution. It is essential for the affiliated priority optimisation system, moreover, that the regulation system additionally be in a position to signalise failure events in respect of vehicles and track installations outwards so as to be able to respond in suitable manner. It is possible to assign each failure event to a unique event category to this end. It is necessary in the process to synchronise the regulation system's very detailed data management capability with the significantly coarser modelling of the priority optimisation system.

The central purpose of the regulation system, however, relates to detecting and resolving conflicts on the basis of the priorities conveyed to it. There are three different strategies that can conceivably be adopted to resolve conflicts on a priority basis, those being manual computer-aided traffic regulation or else automated approaches of either a centralised or decentralised nature. In the case of manual computer-aided regulation, advice filtering, running-time computation and conflict detection are automated whereas actual resolution of conflicts remains the domain of the operative. With centralised automated regulation, conflict automation, too, is performed automatically by a central unit. Decentralised automated regulation, finally, involves the tasks of running-time forecasting and the detection and resolution of conflicts being assumed by several local rescheduling computers networked with one another.

5 Key requirements for the monitoring system

The monitoring system serves to rate traffic processed with the aid of the priority-based regulation system. Rating is effected on the basis of a predefined performance regime co-administered by the monitoring system. The monitoring system operates in the background and is primarily tasked with evaluating traffic notices received from the outside world, in support and independently of operational activities, to the specifications set forth in the performance regime. In the light of this evaluation, the monitoring system detects firmly preset penalty events implying a potential need for a longer-term adaptation of priorities. These events, which are each assigned to a unique event category, are then notified to the priority optimisation system.

Under the performance regime, predefined deviations from the schedule (in particular, delays and lost connections) are numerically weighted by means of penalties. Hence the value of a railway's service offerings and performance will be expressed in terms of transport economics. Reducing the penalties incurred will become a target of operating practice. This is the focus of interest towards which the according of priorities and thus regulation itself must be directed.

Penalties can feasibly be levied relative to the delay arising at a predefined reference point (as a function of delay minutes) or relative to the holding of a given connection.



6 Principal aspects of a priority optimisation system

6.1 Purpose of the priority optimisation system

The priority optimisation system acts to internally optimise priorities in regulation as well as to supply these priorities to the regulation system. To this end, the priority optimisation system obtains failure advice from the regulation system and evaluation advice (penalty payments) from the monitoring system.

Internally, the priority optimisation system distinguishes between “standard” and “degraded mode” or “disruption” priorities (see fig. 6). Standard priorities are the train and connectional priorities, firmly defined for each point in time and for each location, for use in the event of traffic being free of disruptions or only slightly disrupted. Disruption priorities, by turn, involve predefined deviations from “standard” priorities in respect of specified major degraded-mode scenarios signalised by the regulation system (“emergency plans”). In the event of traffic being largely free of disruptions, it is the respective standard priority that is conveyed to the regulation system, otherwise it is the disruption priority defined for the disruption incident concerned that is conveyed.

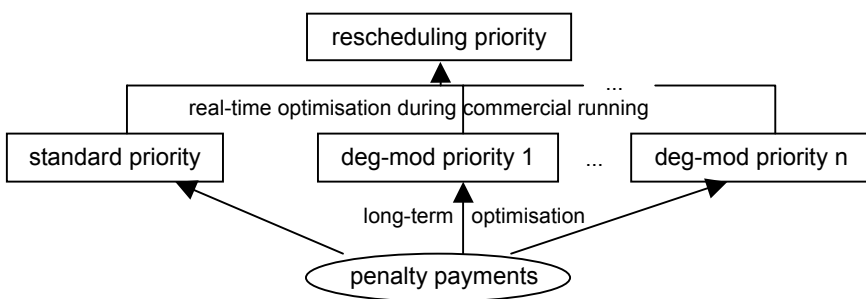


Figure 6: Priority optimisation.

Three possible kinds of change of priority controlled by a priority optimisation system are conceivable in this connection, specifically a scheduled change in the course of regulation, a change on account of a degraded-mode scenario in the course of regulation and, finally, a longer-term adaptation of internal standard and disruption priorities.

Scheduled changes of priority can arise where train priorities are concerned when a train passes from one priority section to another. In such an instance, the regulation system replaces the standard or degraded-mode priority applicable for the former section with the standard or degraded-mode priority applicable for the new section. A change of priority on account of a degraded-mode scenario, by contrast, corresponds to a short-notice temporary switching of rescheduling priorities between standard and degraded-mode priorities or between differing degraded-mode priorities owing to a sudden degraded-mode incident. The switching of priorities in respect of a degraded-mode scenario is determined

beforehand by means of the degraded-mode priorities for train and connectional priorities. It is ensured in this way that there is no need to carry out extensive, time-critical ad-hoc calculations or optimisation work “with wheels rolling”. Where longer-term priority adaptations are concerned, on the other hand, it is the underlying (internal) standard and degraded-mode priority data used in specific instances themselves that are optimised and not the rescheduling priorities. This is effected by means of an evaluation of actual operations by the monitoring system, observing the values prescribed by the performance regime, that is not carried out in real time and is hence largely non time-critical.

In what follows, a practical proposal for putting changes of priority to effect with the aid of active and deductive database technology is made.

6.2 Effectuation with the aid of active and deductive database technology

Recourse can be had to existing database-technology concepts for the purpose of putting a priority optimisation system into practice. Interest centres in this respect on deductive, active and supplementary normative constituents of database management systems (DBMS). Deductive DBMSs allow new deducible facts to be specified, administered and specially derived from explicitly introduced facts, cf. [7, 8]. An active DBMS allows the definition of reactions to be automatically initiated by the DBMS in response to the detection of specific database-related events, cf. [9]. Additionally, a normative constituent of the DBMS enables fixed integrity constraints to be defined that are required to be kept in each consistent state by the database.

The deductive constituent is in particular used in the context of the explicit definition of combinations having a bearing on the according of priorities. The elementary events transmitted by the traffic-regulation or monitoring system are generally unsuitable for use in priority management on their own. It is usually necessary to specify the event and thus resulting internal status changes more precisely as a function of the environment in which it occurs (constraints) before using it in the priority optimisation system. This can be done using deductive rules having the following simplified base structure:

$$\text{combination } (p_1, p_2) \leftarrow \text{status } (p_1, p_2, p_3), \text{ constraint } (p_3, p_4).$$

The active constituent is needed to execute the responses of the priority optimisation system (priority manipulations). These responses are to specific events. These events correspond with insertions or deletions of status information from the traffic-regulation or monitoring system or out of it deductively defined combinations. The active rules have the following simplified base structure:

<i>ON insert</i> (<i>combination</i> (p_1, p_2))	[Event]
<i>IF status</i> (p_2, p_3)	[Condition]
<i>DO change standard priority</i> (p_1, p_3).	[Action]

By support of the normative constituent it is possible to define general fixed regulations of relevance to the according of priorities for regulation in the



relevant network (e.g. statutory precedence provisions, waiting time instructions etc.) with the aid of normative rules or integrity constraints. It is essential that the latter be adhered to at all times regardless of any optimum levels specified under the performance regime. A normative rule contains details to this end of parameters covered by the normative process together with a range of values cleared for the priority with the appropriate normative credentials.

Clearly, it would also in principle be possible to implement a priority optimisation system with the aid of a suitable declarative programming language and hence do without these DBMS constituents. The advantage of the approach delineated here, however, is that the input for implementation and, above all, for future servicing can be significantly reduced given an altered rules scenario. Instead of having to laboriously process the source code, it is merely necessary to adapt the rules data to the database interface.

7 Conclusion

This paper sets out the development of a rules-based system of traffic regulation for railway operations involving priorities optimised on the basis of rules. The priority optimisation system delineated serves an interfacing regulation system as a decision support tool in the event of conflicts arising and operates in detachment from the actual conflict resolution process. With the aid of the priorities prescribed, the interfacing regulation system is able to independently identify potential forms of concrete conflict resolution and put them into practice. It is possible using an affiliated monitoring system to rate regulated traffic on the basis of a performance regime. This provides the wherewithal for further optimising priorities. Recourse can be had to existing deductive, active and normative constituents of a database management system for the core processes of managing and optimising priorities.

The approach portrayed pursues the goal of combining the benefits of optimisation and heuristic procedures for train service regulation into an all-embracing concept. The configurational concept presented here is currently being implemented and tested in prototype form by the Institute of Transport Science at RWTH Aachen University. Further findings are to be anticipated soon therefore.

References

- [1] Pferdmeiges, S. & Schaefer, H., Automatische Konflikterkennung und wissensbasierte Konfliktlösung in der Streckendisposition, *Signal + Draht*, 87, pp. 174-177, 1995.
- [2] Martin, U., Verfahren zur Bewertung von Zug- und Rangierfahrten bei der Disposition, *Eisenbahntechnische Rundschau*, 45, pp. 582-583, 1996.
- [3] Oh, S. M., Hong, S. H. & Choi, I. C., Railway conflict detection and resolution in the Korean railway system, *Computers in Railways IX*, pp. 675-684, 2004.



- [4] Kosemund, M., *Beitrag zum Einsatz von Prozeßrechnern für die optimale Zuglaufregelung auf Eisenbahnstrecken*, Dissertation, HfV "Friedrich List", Dresden, Fakultät für Technik und Naturwissenschaften, 1975.
- [5] Jacobs, J., Reducing delays by means of computer-aided 'on-the-spot' rescheduling, *Computers in Railways IX*, pp. 603-612, 2004.
- [6] Wunderlich, W. & Kockelkorn, E., Bessere Disposition durch schnellere Information, *Eisenbahntechnische Rundschau*, 24, pp. 283-294, 1975.
- [7] Ullman, J. D., *Principles of Database and Knowledge-Base Systems*, Computer Science Press, 1988/89.
- [8] Bry, F., Manthey, R. & Schütz, H., Deduktive Datenbanken, *Künstliche Intelligenz*, 3/96, pp. 17-23, 1996.
- [9] Widom, J. & Ceri, S., Introduction to Active Database Systems (Chapter 1), *Active Database Systems: Triggers and Rules For Advanced Database Processing*, Morgan Kaufmann, pp. 1-41, 1996.



This page intentionally left blank

Simulation of traffic management with FRISO

A. D. Middelkoop¹ & L. Loeve²

¹*ProRail Strategy and Innovation, The Netherlands*

²*Incontrol Enterprise Dynamics, The Netherlands*

Abstract

Due to the high occupancy rate of the Dutch rail network it is difficult to handle future transport demand by the introduction of more trains. The main challenges are on one hand generating a feasible timetable and at the same time meeting performance criteria e.g. punctuality and train delays. Possible solutions are extension of infrastructure capacity and/or adjustment of the timetable structure. Besides those more common measures there is a third option that focuses on traffic management. These measures concern operational and planning issues such as increasing the level of flexibility in daily operation, e.g. by better decision support systems, and decreasing the level of detail in the planning stage. In order to quantify the improvements of such measures, Dutch Inframanager ProRail uses simulation. This simulation approach requires new functionality, regarding train control mechanisms.

One of the first studies performed with this kind of simulation is the important bottleneck Schiphol station. Following this approach the development of a more general simulation tool has been started. This tool is called FRISO: Flexible Rail Infrastructure Simulation of Operations. Important features of this tool are the automatized construction of a simulation model by using a connection to an existing infrastructure database, flexible infrastructure editor including generator functions and the possibility to perform single and multiple (stochastic) simulation experiments.

A special feature is the possibility to connect to an existing *external* traffic management system (TMS). By this connection the effects of TMS-variants may be quantified. The software architecture contains synchronisation mechanisms that allow consistent time management and interaction between the simulator FRISO and TMS or other applications. TMS can take over internal FRISO train dispatching tasks to optimise the train traffic in a local area. It is designed to improve traffic control by means of advisory speeds, routes and order of trains.

Keywords: simulation, traffic management, distributed simulation.



1 Introduction

This paper describes the development of a new train simulator called FRISO, which stands for Flexible Rail Infra Simulation of Operations. After a short introduction of the Dutch Rail Network in section 2, it explains the programme of Dynamic Traffic Management (DTM) in section 3. This programme consists of a number of measures which change the planning and the operation principles of the railway traffic. A particular part of this programme is the development and introduction of a Traffic Management System (TMS) which optimises traffic performance on punctuality, energy consumption and throughput.

Next, section 4 describes the simulation approach to study the effects of DTM, starting with the first experiments/studies regarding Schiphol, the railway station of Amsterdam Airport. Then follows a more general approach, which requires a flexible simulation tool (FRISO) for investigating traffic control solutions for potential bottlenecks in an efficient way. Then section 5 explains a special feature of the approach, the connection of TMS and FRISO as part of a High Level Architecture (HLA), which enables consistent synchronisation of so called simulation federates and flexible exchange of information and functionality/software.

The paper ends with some remarks on current status and following activities.

2 The Dutch Rail network

The size of the infrastructure of the Dutch railway network is moderate compared to most other European countries. The length is 2800 km and it contains about 6500 km of tracks. Extension of the network is foreseen to be ready in 2007. Then a new high speed line Amsterdam-Schiphol-Rotterdam-Belgium, a new connection between the harbour of Rotterdam and Germany (Betuweroute) and a new extra double track connection between Amsterdam and Utrecht will be available.

The main part of the traffic (128 million trainkm in 2002) is used for passenger transport. On a daily basis some 5000 trains carry 1 million passengers. Freight transport is growing more important, but is still a minor part (about 8% of the trainkm performance). Each day about 280 cargo trains transport 80 kton of freight.

The occupation rate of the Dutch rail network is high [1.]. This is illustrated in figure 1. In this graph the trainkm/netkilometer (x1000) rate per year for a number of countries is shown.

The actual state of the railway-infrastructure and the expected growth of passengers- and freight traffic don't match. The current timetable structure cannot transport these amounts of passengers and freight, so more trains are necessary. The introduction of more trains causes not only capacity problems in the network, but also a decrease of the performance, measured by punctuality and train delays. Insufficient capacity on potential bottlenecks may be solved by extending the infrastructure capacity. However, building railway infrastructure is very expensive and it takes a long time to design and fit it into the environment.



So new and innovative measures are necessary to be able to operate an attractive and cost-efficient timetable. One of those measures, a programme called Dynamic Traffic Management, is explained in the next section.

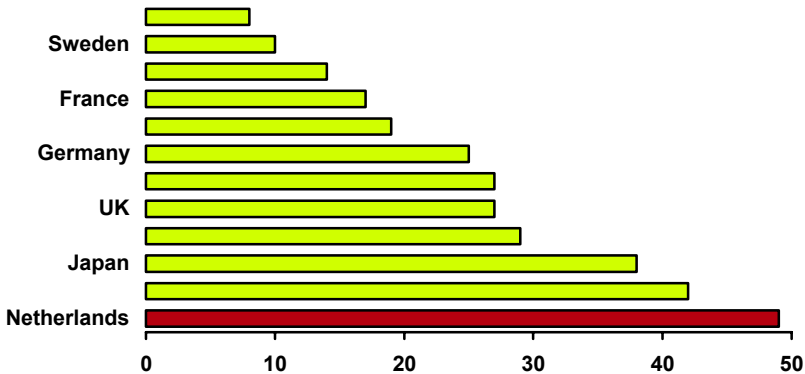


Figure 1: Occupation rate: x1000 trainkm/netkm per year in 2002.

3 Dynamic Traffic Management

At ProRail, the Dutch rail infrastructure manager, there is a programme of new innovative ideas which investigates the introduction of Dynamic Traffic Management (DTM). Main issue of DTM is to accept the fact that in daily operation deviations of the planned situation will occur. The aim is to manage these deviations effectively the moment they take place to improve the performance. The programme consists of measures that change the design and operation of the timetable by introducing innovative planning principles and advanced new control strategies. Problems are not statically avoided in the planning, but dynamically solved in real time. To predict the effect of those measures on timetable quality ProRail uses a simulation approach.

3.1 Changing planning principles

Railway transport has a lower level of operational freedom than other transport modes. Railway companies plan the train services in a very detailed way. The train order at crossings, junctions and platform tracks is fixed more than a year in advance to prevent conflicts and to check the capacity. In practice trains rarely travel exactly to the plan as illustrated in figure 2. In the right part of the figure each line represents the running of a train on a certain day in March 2003. The plan contains fixed recovery times to handle daily variations. The DTM concept aims on a flexible use of these recovery times. Thus the main focus is not on planning but on performance of the train services.

The DTM concept puts the reliability first. An *organic planning* approach takes into account the daily variation performed in practice. The required running



and dwelling times as demonstrated in practice serve as the starting point for constructing a reliable timetable. It determines how much margin time is needed to obtain a certain level of reliability. This feedback can support both the short term operational process as well as the long term planning process.

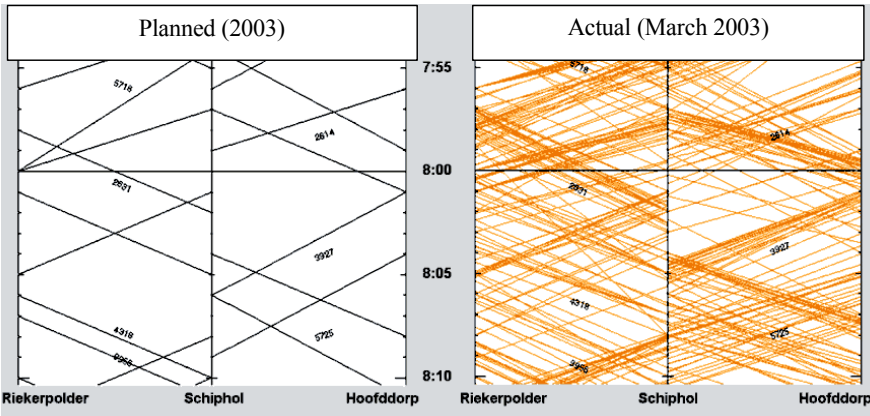


Figure 2: Planned and actual use of the train services around Schiphol station during march 2003 (source: ProRail)

Once we accept that disturbances are part of our reality, we need the flexibility to handle them. Therefore we must avoid certain details in our planning. This can be done in different ways: time windows, dynamic use of tracks, dynamic change of order. *Flexible planning* places a train in a time window (no exact time) and plans it to arrive at a certain range of platform tracks. The precise platform track is decided (and communicated) during operation, as the actual traffic situation is known. The earliest departure time and the latest arrival time are published as a customer timetable. The first bottleneck investigated in this way is described in section 4.1.

Flexible planning also enables an exchange of buffer times. In bottlenecks buffer times are no longer fixed between trains. Some conflicting train movements are thus allowed in the planning. In practice the conflict might not even occur, or a solution will be found. The best order of train movements is chosen real time. This assumes a very good traffic control including powerful decision support systems.

3.2 Supporting the traffic control

As the planning is less detailed the details may be managed in real time. Traffic control thus requires more real time actions. The dispatcher needs more support to perform this extended task. DTM defines two types of support: simple and advanced traffic control.

Simple traffic control uses predefined sets of dispatching rules, based on off-line analysis. Advanced traffic control uses real time control, based on on-line

analysis. Both approaches need support systems either to improve decisions or to automatically perform (parts of) the dispatching task.

Local systems act only “on the spot” and “now”. They may act according to dispatching rules such as “First Come First Served”. In some situations the rules must be extended with detailed local dispatching rules.

Advanced traffic management systems take into account all traffic in a certain area. They predict conflicting train movements, calculate an optimal traffic flow and present advisory speeds to train drivers. They also advise a change of order or route to the dispatcher. In the COMBINE research project [2] a demonstrator has shown the effects of such a real-time traffic management system (TMS). It optimises the traffic performance on punctuality, energy consumption and/or throughput. In the Netherlands a prototype has been tested in operation. The results were promising, punctuality is expected to rise about 3%, the energy consumption about 7%. This is reason to study the effects of the TMS on other potential bottlenecks. To prevent high costs of testing in reality and redevelopment of functionality, the TMS therefore should be connected to a simulator.

This requires a very flexible simulation environment, which must be able to handle different traffic control strategies, perform stochastic simulation experiments and connect easily to different versions of traffic management systems.

To enable future exchange of new components and already existing functionality both the simulator and the TMS will be part of a High Level Architecture (HLA). This HLA [3] supports interoperability between different simulation models and enables the reuse of existing functionality and software. This is further elaborated in section 4.2.

4 Simulation approach

To predict the change in the performance of a timetable, as a result of the introduction of simple or advanced DTM-measures, a simulation approach is required. One of the first bottlenecks in the network where DTM-solutions are necessary is the railway station of Amsterdam Airport, Schiphol. Section 4.1 describes a simulation study, where this approach is applied, supported by a prototype version of a simulator. The results were reason to investigate further introduction of DTM-principles and to start the development of a generic flexible simulation tool called FRISO. This tool is described in section 4.2.

4.1 Schiphol simulation study

The railway station of Amsterdam Airport, called Schiphol, is situated underneath the entrance hall of the airport, so travellers can enter and leave the airport by train very easily. The layout of the station is shown in figure 3.

Today 14 trains per hour run through the tunnel, but in less than 4 years time this should be almost doubled. Given the current rail infrastructure and the origin and destinations of the train services, the number of conflicting situations



between trains is expected to increase rapidly, which will affect train delays and punctuality. Adding tracks is extremely expensive due to the underground location and changes to the timetable do not seem to be sufficient, so other measures are necessary.

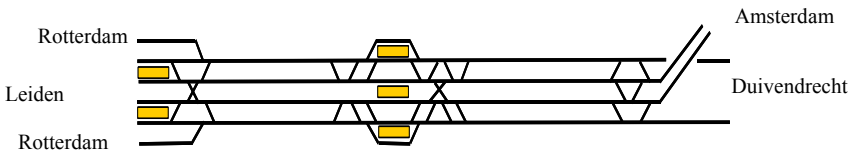


Figure 3: Layout infrastructure Schiphol.

Therefore combinations of small changes to the infrastructure and DTM-solutions were evaluated. The first proposed solution concerns decreasing the minimum time between subsequent trains. By removing some switches, it is possible to shift signals towards the platform. In this way, it takes less time for waiting trains to enter the platform. A second solution for decreasing the minimum time between trains is to decrease the setting times of signals significantly.

As a third solution based on DTM, introduces time windows into the timetable. Trains can leave stations earlier due to the less stringent departure times, thus optimising dwelling times. The introduction of a control strategy, based on the First Come First Served principle is a fourth option. The order of trains as given in the plan is not preserved but based on the actual arrival times of trains at a controlled area. As final solution, the dynamic use of platform at Schiphol is proposed, under the condition that passengers should be informed in time of platform changes.

The above mentioned solutions were combined in several scenarios. To determine the impact of the solutions on resulting running times simulation was used. Especially for the Schiphol area a prototype of a flexible simulation tool was implemented. This prototype generates simulation models out of existing databases and is designed in such a way that it is very easy to change control strategies, setting times of signals and platform usage. As performance indicator the 87 percentile of running times was measured. The number of trains per hour was increased at each experiment, resulting in longer running times. The current planned running times of trains in the study area was compared with the experiment results.

The simulation study showed that without changes in infrastructure, the desired amount of trains is possible when planned running times are extended with 3 minutes. Changes in infrastructure have a positive effect on the capacity, but not enough to handle the desired amount of trains with the current planned running times. DTM-measures are therefore necessary. The most effective scenario contained first come first served principles together with dynamic platform usage. However, to satisfy a desired punctuality of 87% 1 minute extra running time is needed in the timetable. The results were so promising that at the



end of 2005 a real life test was performed at Schiphol, showing that the simulated control strategies are also feasible in practice.

The results were reason to test DTM-principles in reality as a pilot project and to investigate introduction of both simple and advanced DTM-principles on other bottlenecks. Necessary further development of the simulation prototype to a flexible tool is described in the next section.

4.2 FRISO: Flexible Rail Infra Simulation of Operations

Performing a simulation study is very powerful but has also some disadvantages. These are related to the time effort needed, to the software architecture and therefore the costs.

Looking at the time needed for a simulation analysis, a great part of it is used for building (often manually) and validating the simulation model of a situation. The investigation of a new situation causes an equal development effort. Here a generation of the simulation model based on input from company databases, in combination with smart editors for making timetable and infrastructure variants, can reduce the time effort significantly.

Often railway simulators are tailor made for a certain railway system. They originate from the time that object oriented programming was not yet a standard. This makes it difficult and expensive to introduce properties of new other railway systems or to change existing properties, e.g. regarding the way the signalling is modeled.

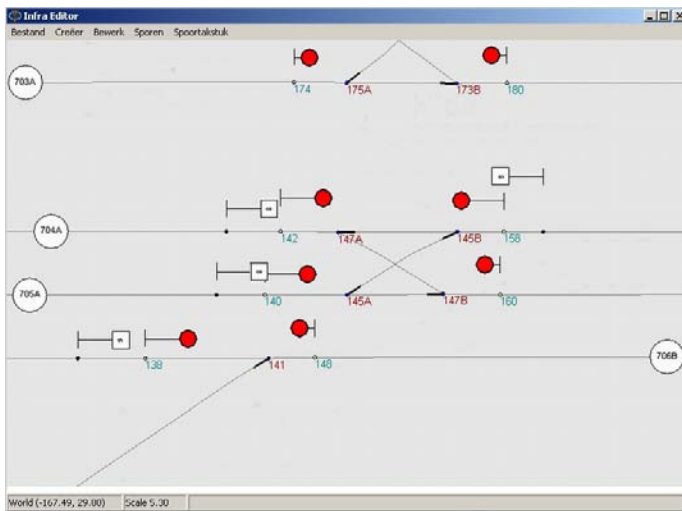


Figure 4: FRISO infrastructure editor.

With the development of FRISO, ProRail adopted the approach of SIMONE, a countrywide network simulator [4], to overcome these disadvantages. Both tools are based on a general simulation language/tool called Enterprise



Dynamics, which is used in a lot of industries. On top of the simulation engine ProRail developed a library of railway components. A simulation model may be generated automatically by ‘filling’ the components with timetable and infrastructure data. The user can spend most of his time on the definition of simulation experiments and the analysis of the results.

FRISO models the following elements of the railway infrastructure: track layout, signalling system, route setting, and interlocking, see figure 4. In the timetable trains are allocated to routes. The data is imported from timetable and infrastructure databases (Infra Atlas). To make simulation variants FRISO has a number of edit functions, to change infrastructure, timetable and traffic control options. These functions will be enhanced in future versions with smart generators for signalling and interlocking elements and route definition.

A special feature of the flexibility is the possibility to connect to other applications that may take over parts of the functionality. A first example, the connection between an advanced traffic management system and FRISO, is explained in section 5.

5 FRISO and TMS

A train simulator has functionality, mostly a set of rules, to control the train traffic. For testing the TMS benefits these rules become partly redundant. The idea is that the TMS will improve the traffic performance by optimising the situation on punctuality, energy consumption and/or throughput. It therefore must be able to control the trains and exchange information on the actual position and status of the trains with the simulator. Connecting existing functionality as the TMS to the simulator is a form of distributed simulation. The simulation scenario will contain several components, so there is a need for a control mechanism. Especially the time management is an important issue. The FRISO-TMS combination uses an architecture which enables interoperability of simulation models, the High Level Architecture (HLA) [3]. This standard promotes re-use of (different kinds of) simulation models and their components. It specifies interfaces between components and defines the steps for the development and execution of a simulation scenario. The applications in such a scenario are also known as federates. They form a federation and have to cooperate, communicate and synchronize mutual information. In the simulation FRISO informs the TMS about location and status of the trains, occupation of the infrastructure, allowed speed profile restricted by the signalling system and allocation of routes. The TMS returns advisory speeds, booking of routes and (later also) change of track allocation.

The architecture of an HLA-connection between TMS and FRISO is shown in figure 5. To translate messages of FRISO and TMS to HLA, mappers are used. These are connected to the run time infrastructure based on HLA principles. HLA takes care of time synchronization, saving and restoring simulations and updating information. By using the mappers and HLA it becomes very easy to replace the TMS with another traffic management system or to replace FRISO



with real life trains. The only modification needed then is to implement a mapper for the newly connected system, translating messages into HLA-messages.

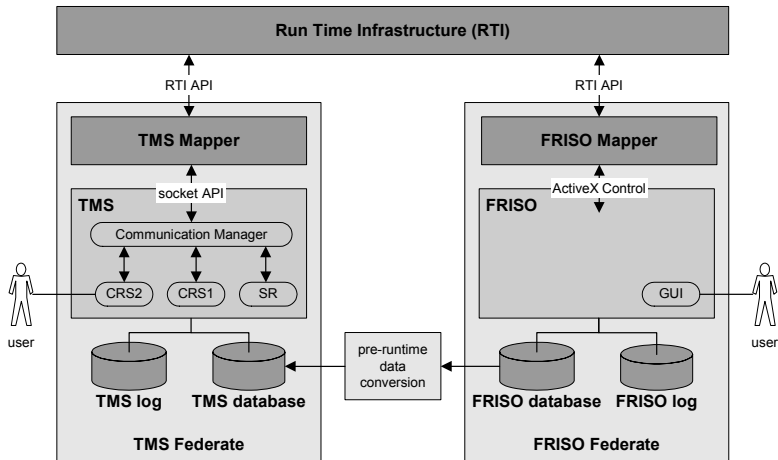


Figure 5: Architecture FRISO and TMS.

6 Future research

The current state of the development is that the first version of the simulator FRISO is validated on a part of the network. The next steps of the development will increase the flexibility of the tool by improving the user interface and by introducing more input generators.

The connection with the TMS is ready. It will be used in a series of simulation experiments to investigate the DTM-effects and the tuning of the TMS on several junctions. First results will be available in June 2006.

References

- [1] Poort, J.P.: Limits on utilization (in Dutch). NYFER (2002)
- [2] De Vries, D.: COMBINE a decision support system for real time traffic control. Computers in Railways VIII (2002) 1003–1010
- [3] IEEE Standards 1516.3: Recommended practice for High Level Architecture (HLA), Federation Development and Execution Process (FEDEP). Institute for Electric and Electronic Engineers (2003)
- [4] Bouwman, M., Middelkoop, A.D.: Testing the stability of the rail network. Computers in Railways VIII (2002) 995–1002



This page intentionally left blank

Influences of station length and inter-station distance on delays and delay propagation on single-track lines with regional rail traffic

O. Lindfeldt

Division of Traffic and Logistics,

*Department of Transport and Economics, Royal Institute of Technology,
Stockholm, Sweden*

Abstract

Train services on single-track lines suffer from time losses due to crossings, imposed by the bidirectional traffic. The time losses are caused by constraints in the infrastructure and delay propagation, which give a stochastic contribution that varies from one crossing situation to the other.

Two examples of infrastructure improvements that decrease the time loss are examined: increased track length at timetabled crossing stations and decreased inter-station distances. A mathematical model is used to evaluate these improvements.

Longer station tracks seem to be very efficient when traffic intensity and delay variances are moderate. Shortened inter-station distances give less effect but are less sensitive to delay variance and give valuable additional line capacity.

The model used assumes independence between crossing trains, which imposes a moderate capacity utilisation. In more congested situations simulation methods are needed to make more complex crossing patterns possible.

Keywords: single-track, delay propagation, partial twin-track, inter-station distance.

1 Introduction

On single-track lines train services often have lower average speed and worse punctuality than on twin-track lines. These facts are caused by the bidirectional traffic that calls for crossings. These crossings take place on crossing stations with two or more parallel tracks. In many cases the crossing itself leads to time



losses, compared to a crossing on a twin-track line. These time losses consists of three parts:

- A static lowest time loss that is given by the infrastructure. This is a minimal (theoretical) time loss that occurs also when the crossing trains arrive at the station in an optimal way.
- A planned time loss that is due to infrastructure imperfections, such as unequal inter-station distances, irregular traffic or other types of planned time supplements.
- A stochastic time loss due to arrival delays, varying from one crossing situation to the other.

These three factors are not additive. High values on the planned time loss can overtake the static one and make the local infrastructure less constraining. A planned time loss also decreases the impact of arrival delays. One important case is when a crossing is combined with an ordinary stop for passenger exchange. In such situations influences of infrastructure constraints are often minimized and the impact of arrival delays decreased. Assuming perfect conditions in surrounding infrastructure, no time supplements and no stops for passenger exchange, the static lowest time loss and the stochastic time loss due to crossing can be analysed systematically.

Regional train services have some important demand features that are hard to combine with constraints imposed by single-track lines:

- Low travel times. Regional train services must reach a high average speed.
- High frequency. At least in morning- and afternoon rush hour a regular and frequent service is required.
- High punctuality. People travelling to and from their work places require a high punctuality.

Trying to meet these requirements, crossing restrictions is a great problem. This paper describes two examples of improvements that decrease time loss due to crossing: longer station tracks and shorter inter-station distances. Other measurements, such as higher speed in points etc, will be left to future research.

2 Method

2.1 Model

In order to analyse crossing phenomena on single-track lines, a mathematical model has been constructed. In the model it is possible to vary several parameters, hence different kinds of analyses can be performed. Most important variables are: inter-station distances, track lengths for each station, speed restrictions at points, interlocking time for each station, vehicle retardation and acceleration data and distributions of arrival delays (primary delays).

One advantage of this model is that one or more variables can be varied systematically. As shown later, characteristics of different infrastructure designs can be evaluated very easily.



In order to make the model as simple and transparent as possible, some idealizing assumptions were made:

- Signalling system: ERTMS level 2, e.g. continuous update of driving permissions.
- Crossing trains are independent before crossing. This assumption requires a moderate capacity utilisation. At higher levels, interferences between trains will cause further restrictions on crossing. The model serves to give a first idea about characteristics of single-track lines. A very interesting analysis of highly congested cases will be performed by means of a simulation later on, in order to find limitations of this simple model.
- Everything that happens outside the model is taken into account by distributions of arrival delays. Each train only cross one other train in the modelled line.
- No time supplements in timetable. One output of the analysis is the need for (and effect) of time supplements.
- Dispatching follows the very simple rule to minimize the sum of added delay for crossing trains. This assumption also requires a moderate capacity utilisation.
- No gradients.

Given the infrastructure model, the vehicle model and the idealizing assumptions delay characteristics for a line can be calculated. Fig. 1 shows how the additional delay, due to crossing, depends on the difference in arrival delay.

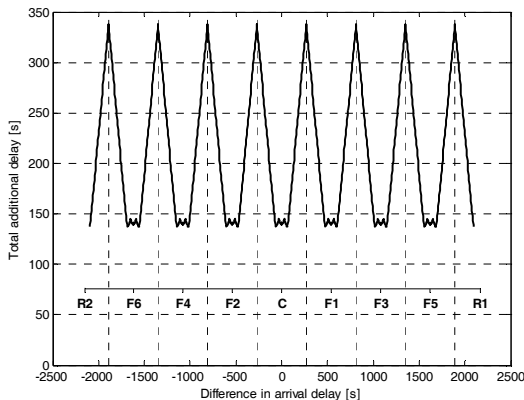


Figure 1: Total additional delay as a function of difference in arrival delay. Each crossing station is used during a time interval between two adjacent maxima (vertical dotted lines). Station signatures R2-R1 shown. C denotes timetabled crossing station.

For every five seconds of delay difference the model has solved a crossing conflict. This is done by first detecting the conflict and alternative crossing solutions. Second a crossing course, involving train movements, is performed for



each station alternative. Finally the station giving lowest total additional delay is chosen to give the delay value for the current delay difference.

Note that the reference level, additional delay = 0, represents a situation where both trains run the line without any time losses due to crossing, which corresponds to a twin-track line. In fig. 2 the tested line is schematically shown.

The length of the model, 120+20 km, is chosen statistically so that at least 99% of all crossings will take place within the model, given that both distributions of arrival delays follow a specific worst-case distribution with very high variance, taken from the Swedish railway's delay data.

In solving crossing conflicts, the model has to choose the best station for each conflict. Therefore the border stations, B1 and B2 in fig. 2, must be treated carefully, since the model does not know anything about stations beyond these.

So far no attention has been paid to the fact that the difference in arrival delay follows a distribution. Thus the outcome of different differences of arrival delays will be given different weights according to a distribution. This distribution can be calculated through convolution of the two crossing trains' arrival delay distributions:

$$f_Z(z) = \int f_X(\tau) f_Y(\tau - t) d\tau. \quad (1)$$

In eqn (1) τ denotes arrival delay for the down train, t denotes arrival delay for up train and z denotes difference in arrival delay. f_X and f_Y are the corresponding probability density functions and $f_Z(z)$ the resulting probability density function. Eqn (1) requires that the delays of two crossing trains are independent.

As a first step, negative exponential distributions have been used. Real delay statistics show that expected values on single-track lines in Sweden are 150-350 seconds. A negative exponential distribution, given the same expected value, shows a higher variance than empirical distributions. Later in the research project, other, more sophisticated distributions will be examined in order to get less discrepancy.

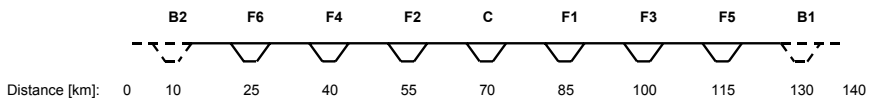


Figure 2: Example of line model with uniform inter-station distance 15 km and track length 750 m (Swedish standard).

Two equal negative exponential distributions ($f_X = f_Y$) give a Laplace distribution when convoluted, eqn (2). In fig. 3 this distribution is shown for three values on the scale parameter λ . Corresponding expected values for f_X and f_Y were 150, 250 and 350 s respectively. Note that the same distribution has been used for down- and up trains, making the curves symmetrical.

$$f_Z(z) = \frac{\lambda}{2} e^{-\lambda|z|}. \quad (2)$$



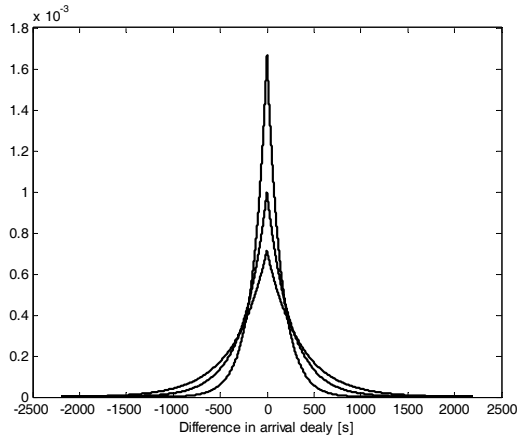


Figure 3: Three levels of probability density functions for difference in arrival delay. Standard deviation 210, 350 and 475 s.

A very interesting question is how sensitive the outcome of different infrastructure improvements is for the shape of this distribution. Some measures will turn out to be quite dependent, while others will not.

2.2 Choice of variables

Several improvements are possible in order to get less time consuming crossings and lower disturbance sensitivity. Two interesting alternatives are:

- Increased track lengths at the station where the crossings are planned to take place. When station tracks are longer than a certain length, crossings without any interference between trains are possible. To make this kind of improvement cost efficient, some kind of regularity with a basic interval timetable is required so that utilisation is high. This timetable also has to be stable for several years. A severe drawback for these long stations is just that the stations on the line have different performance. Choosing another timetable, whose crossing pattern differs from the optimal one, results in low infrastructure utilisation.
- Decreased inter-station distances. If inter-station distances are decreased the maximum additional delay will be lower, but the minimum additional delay will not change. An important advantage of decreased inter-station distances is that it increases the overall capacity of the line. This means that the investment can be used either to run more trains, maintaining existing punctuality, or to run the same number of trains as before the improvement but with a higher punctuality.

To evaluate the effect of increasing track length at the timetabled crossing station, C in fig. 1 and 2, track length was increased stepwise by 500 m from 750 m to 28 250 m. Thus inter-station distances F2-C and C-F1 decreased by 250 m each time. For moderate increases (≤ 15 km) this method seems to be of



practical interest. The shortened inter-station distances on each side of station C is then part of the total improvement. For longer extensions the question is whether to adjust flanking inter-station distances for capacity or economical reasons.

The other variable of special interest, inter-station distance, requires a different approach. In this case all stations, C included, were of same length: 750 m. An interval of interesting inter-station distances was defined as 3-15 km. Between these limits calculations has been performed with a step of 500 m. As already mentioned, a minimum model length of 120+20 km was set up using delay statistics. Given this minimum, the number of modelled stations increased with decreasing inter-station distances. Moreover, the number of stations had to be odd (symmetry reasons) and so this theoretical number, $120/\text{inter-station distance}$, must be rounded upwards to nearest odd integer.

These adjustments result in different total model length for different inter-station distances, all of them ≥ 140 km. The varying model length could make comparisons more difficult, but since the probability of crossing outside the middle-120 km model part is negligible, this will not cause any real problems.

3 Results

3.1 Increased track length of timetabled crossing station

Fig. 4 shows how the total additional delay depends on the difference in arrival delay. The upper, solid, curve represents a reference case with 15 km inter-station distances and 750 m track length for all stations (shown in fig. 2). Note that each station, located between two adjacent maxima, has a centred interval with low additional delay of about 140 s. At these intervals both trains arrives at the station so simultaneously that none of them has to brake all way to stop. On the slopes up towards a maximum, where the crossing is moved to the adjacent station, one train has to stop and wait, which gives linearity. On the slopes stochastic time loss is dominating, whereas in the minimum areas infrastructure constraints, such as speed restrictions in points and signal interference, due to short station tracks, is the dominating cause of additional delay.

If track length increases, the curve for total additional delay will shift in the time interval covered by station C. This is shown in fig. 4. Equidistance between two dotted curves is 4 km, so that the upper one represents a track length of 4 750 m and the lower 24 750 m. Due to all symmetries in the model parameters the result is quite easy to understand:

- Maxima surrounding the extended station get lower. This is a result of shortened inter-station distances on each side. Note that dotted curves coincide with the reference curve out to the left, and right, of intersection at the new maxima. This means that an extension only affects a specific (local) time interval.
- Curves corresponding to longer station tracks get a new, lower and wider minimum (90 s). On this minimum level time loss depends only on infrastructure constraints, e.g. speed restrictions at entrance and exit



points for one of the crossing trains. Within a time interval crossing can take place without signal interference between the trains. The length of this interval is equal to the width of the minimum area, which depends on the track length.

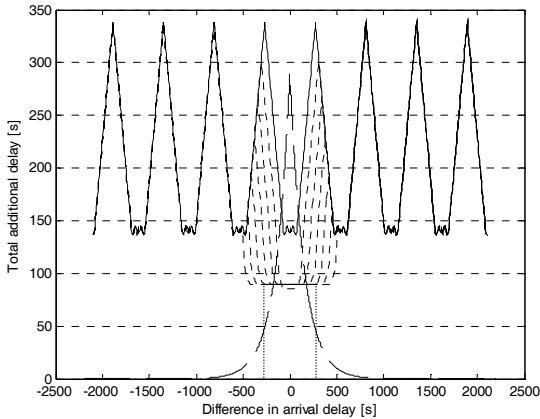


Figure 4: Total additional delay as a function of difference in arrival delay.

Fig. 4 also shows the probability density function for difference in arrival delay (standard deviation 210 s), dashed line reaching 0-level at $\pm 1\,000$ s. Note that a track length of 16 750 m covers a big share of the area below the density curve (between vertical lines). The effect of an extension is thus double. First the minimum level becomes lower, leaving only static infrastructure factors. Second the interval within which the minimum value occurs becomes wider. A greater share of crossings gets the minimum additional delay. These facts result in a strong decrease of average additional delay for the first kilometres of extension.

3.2 Decreased inter-station distances

The result of different inter-station distances is shown in fig. 5. The upper, solid curve still represents a reference case where all inter-station distances are set to 15 km and track length to 750 m. A reduction to 9 km inter-station distances gives a second curve, dotted; with maximum additional delay about 225 s. A further reduction to 3 km gives a curve that oscillates around 140 s, hard to distinguish in the figure.

The effect of decreasing inter-station distances is lower maxima and an increasing share that lies at minimum levels (140 s in this case). In fact, the decrease in maxima is linear from 15 km to ca 5 km. In this interval stochastic delay propagation plays an important role to build up additional delay, although the static time loss is important in situations where trains arrive simultaneously to a station. The stochastic delay propagation is proportional to inter-station distance.



At shorter distance, 3-4 km, the static time loss due to infrastructure constraints dominates and covers stochastic effects. Thus stochastic effects cannot be seen in fig. 5 for inter-station distance 3 km.

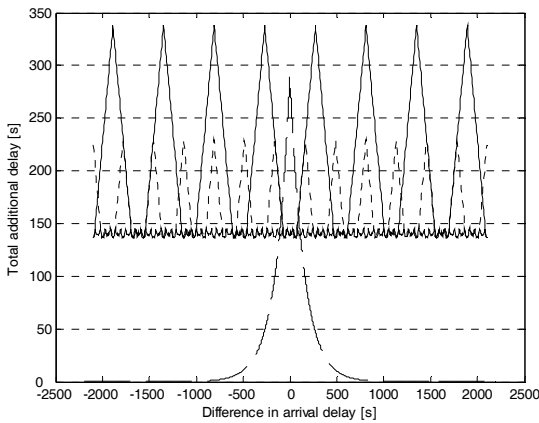


Figure 5: Total additional delay as a function of difference in arrival delay.

3.3 Average total additional delay and standard deviation

The average total additional delay is given by eqn 3:

$$\overline{TAD} = \int f_z(z) * g(z) dz. \quad (3)$$

Here z denotes difference in arrival delay, f_z is the probability density function and g is the total additional delay function shown in fig. 4 and 5. Curves for average total additional delay are shown in fig. 6. The results of three different f_z -functions, standard deviation 210, 350 and 475 s, are shown.

In fig. 6 solid curves represent average total additional delay and dashed curves standard deviations for total additional delay. In the left part, showing results for increased station length, average of total additional delay falls from a maximal level of about 200 s, consisting of static time loss and a great part stochastic delay propagation, towards a lowest static time loss (here 90 s) where delay propagation is almost negligible. Note the effect of higher variance in difference of arrival delay that makes these curves start at a higher level and fall slower.

For the inter-station distance case the minimal total additional delay depends on the fact that these “short” stations always give time losses due to signal interference between crossing trains. Therefore these curves (right part of fig. 6) fall from maximal level towards this static lower limit determined by infrastructure performance. Note that a change in variance for the difference in arrival delay does not influence the average delay very much.



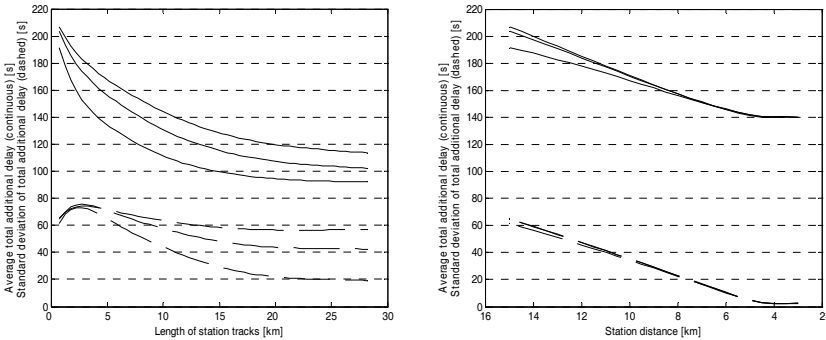


Figure 6: Average total additional delay (solid) and standard deviations of total additional delay (dashed) for increasing track length, left part, and decreasing inter-station distances, right part. In each case results from three different distributions (fig. 3) for difference in arrival delay are shown.

4 Conclusions

Presented results show some principle differences between the two alternative improvements.

Extended track lengths are most efficient when variance in arrival delay difference is low. A lower standard deviation of arrival delay difference than 210 s, lowest curves in fig. 6, and thereby a higher precision, would make this improvement even more efficient. Still higher effect could be obtained with higher speeds in entrance and exit points at the extended station, resulting in lower minimum time loss. Shortened inter-station distances, especially values under 10 km, seem to give a result that is more independent of variance of arrival delay difference. Both these results provide a moderate capacity utilisation, so that alternative crossing stations are available in most frequent crossing situations.

Extended track lengths only have a local effect. Rather high minimum levels of standard deviations reflect this. Crossing situations that fall outside the extended station still result in high additional delay due to long inter-station distances. In case of shorter inter-station distances standard deviations fall faster and to a much lower minimum level.

Low variance in arrival delay difference is an important requirement to obtain independence between crossing trains. Therefore it is very interesting to find out at which variance level, and/or capacity utilisation level, this simple model starts to give serious under-estimations of additional delays. This check will be done by simulation in RailSys, which makes interactions between several crossings possible through active dispatching. A qualified guess is that the simple model works quite well if traffic intensity is lower than three trains an hour and direction, provided that variance in arrival delay is moderate.



The studied cases are quite idealized in themselves. Therefore further work will be done to examine asymmetric cases: between traffic directions, in the infrastructure etc. Extended stations are constructed for one (or a few) timetable structure(s). This is a very important (negative) property that can be examined by systematic change of planned crossing position.

A very important situation is when a (planned) crossing is combined with a regular stop for passenger exchange, which gives lower time loss. An analysis of such cases can point out how timetables should be constructed.

Fortunately this work will contribute to a deeper understanding of the features of single-track railway lines and the possibilities to get fast services with high punctuality on such lines. The complete work will be presented in a licentiate thesis in 2007.

Acknowledgements

This project has been financially supported by the Swedish National Rail Administration.

References

- [1] Barter, W., Forecasting robustness of timetables, *Comprail 2004*.
- [2] Fröidh, O., *Introduction of regional high speed trains. A study of effects of the Svealand line on the travel market, travel behaviour and accessibility*, 2003.
- [3] Mattsson, L-G., Ways of deriving train delay relationships, proceeding from the 1st International seminar on railway operations modelling and analysis, Delft, 2005.
- [4] Wendler, E., On waiting times of scheduled train paths in non-cyclic timetables, proceeding from the 1st International seminar on railway operations modelling and analysis, Delft, 2005.



Simulation of disturbances and modelling of expected train passenger delays

A. Landex & O. A. Nielsen
*Centre for Traffic and Transport,
Technical University of Denmark, Denmark*

Abstract

Forecasts of regularity for railway systems have traditionally – if at all – been computed for trains, not for passengers. It has only relatively recently become possible to model and evaluate the actual passenger delays. This paper describes how it is possible to use a passenger regularity model to estimate the actual passenger delays. The combination of the passenger regularity model with railway simulation software is described, demonstrating the possibility of predicting future passenger delays. The described passenger regularity model is run daily to calculate the passenger delays of the Copenhagen suburban rail network the previous day. The results obtained with the passenger regularity model used together with the simulation software are very similar to the daily calculated passenger regularity of the Copenhagen suburban network. As the combined method includes simulation software and reflects the actual passenger regularity, it is possible to use a combination of a passenger regularity model and simulation software to evaluate and compare future scenarios.

Keywords: railway planning, timetable, regularity, simulation, passenger delay.

1 Introduction

Relatively recently has it become possible to model and evaluate the actual passenger delay on large scale railway networks. The method used to model and evaluate actual passenger delays was presented in 2004 by Nielsen [2] and has since been optimised and evaluated [3, 4]. In the planning process, the passenger delays are often calculated by assuming that no passengers transfer to other trains or update/change their route choice when delay or cancellations occur. This assumption does not reflect the passengers' travel behaviour.



The first part of the paper briefly describes how to model passenger delays based on a comparison of realised timetables to planned timetables. A passenger strategy is presented, in which passengers plan their route according to the planned (announced) timetable, but start reconsidering their route within a certain threshold after a delay or cancellation of a train [4]. The route choice model, obtained from the passenger strategy, is run each night to evaluate passenger delays in the Copenhagen suburban rail network the previous day [5] and evaluates the impact of train delays on passengers. The model has shown that, due to delays caused by e.g. passenger/door interactions when timetables are stressed and when trains carries more passengers in the rush hours, passenger delays are greater than train delays [4].

Although the model presented in the first part of the paper is used to evaluate the already run timetable, the model can also be used for planning purposes. The second part of the paper describes how the model can be combined with railway simulation software such as RailSys, making it possible to predict the expected passenger delays for different timetable alternatives. The simulated timetables are exported to the passenger delay model for comparison with the planned timetable.

The last part of the paper demonstrates that a detailed timetable based passenger delay model together with railway simulation software can be used to evaluate different timetables in the planning process. The evaluation can estimate the expected train delays as well as the daily passenger delays. Furthermore, the model can be used to evaluate in which part of the network passenger delays pose problems.

2 Calculating passenger delays

The core idea of the model is modelling passenger delays by assigning a time-space trip matrix on the realised timetable. This is compared to a calculation where passengers were assigned on the planned timetable (the announced official timetable).

It is assumed that passengers plan their optimal desired route according to the planned timetable. If delays occur, exceeding a certain threshold, passengers are assumed to reconsider the route at that point in time and space along the route. If a train is completely cancelled, passengers reconsider their choice without a threshold.

As a benchmark (minimal passenger impact due to the delays), an optimal all-or-nothing route choice model can also be used on the realised timetable. This model assumes passengers to have full knowledge of future delays at the beginning of their trip and to choose optimally in accordance to this knowledge. The difference between the solutions obtained with the two methods (the optimistic and pessimistic) is a measurement of the additional loss of missing passenger information, combined with slow passenger responses to changes in the schedule.



3 Calculating passenger delays by simulation

Calculating passenger delays of the actual performed operation is of interest to evaluate the train company and to identify aspects or routines that could be improved. If it is possible to predict or estimate the future passenger delays, it is possible to evaluate changes in the infrastructure and/or the timetables already when deciding new infrastructure and/or timetables. To evaluate infrastructure changes and timetables it is common to evaluate train delays by simulation. It would thus be obvious and interesting to calculate passenger delays in the same procedure.

To calculate the passenger delays by ordinary railway simulation software such as RailSys, it is necessary to build up the infrastructure and the timetables to be simulated. The rules of operation are then set up together with a set of delay distributions to simulate disturbances in the operation. It is now possible to run a simulation of the train operation with the chosen delay distributions. After running the simulation, it is possible to evaluate the infrastructure and the timetable whereupon improvements can be considered. The work process of the simulations can be seen in figure 1; the arrow describes the workflow.

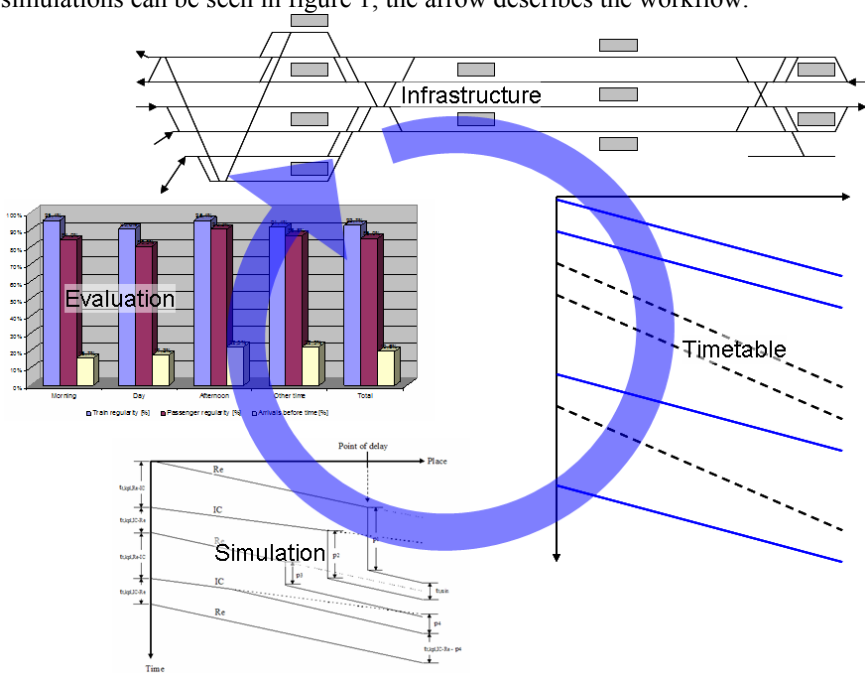


Figure 1: Principles of workflow in rail simulation projects.

Calculation of the passenger delays requires result data from the simulation to contain information of both the planned and all the realized/simulated timetables for all arrivals and departures. The RailSys output file *Fahre++*.pro contains this information. These results must be transferred from the railway simulation



software to the passenger delay model by a simple import-export tool developed in VB.Net.

Calculation of passenger regularity is initiated by coding the infrastructure and creating the timetable. The rules of operation and the set of delay distributions are then defined. To ensure that the model reflect the real life operation, simulations are run and changes made in the rules of operations and the setup of delay distributions. When the model has been calibrated, the simulation is run. It is now possible to evaluate the train delays. However, to evaluate the passenger delays it is necessary to export the simulation data (the *Fahre++.pro* file in RailSys) to the passenger delay model before running the model. The workflow of calculating the passenger delays can be seen in figure 2. The simulation of operation, export to passenger delay model and calculation of passenger delays simulates the impacts of one simulated day of operation. To calibrate the model and to obtain a delay distribution, it is therefore necessary to repeat the third step a number of times before the evaluation.

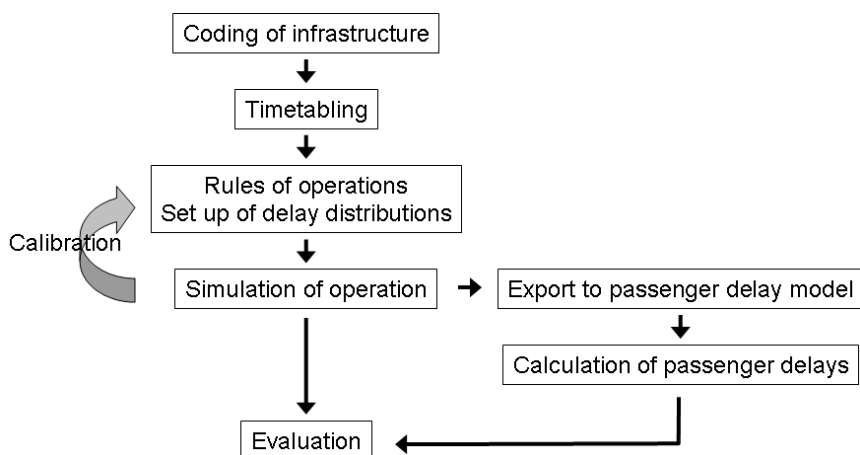


Figure 2: Workflow of simulating disturbances and modelling expected train passenger delays.

4 Simulating disturbances on a large scale network

The entire Copenhagen suburban rail network, including 85 passenger stations, was used for the simulations. The route network contains of 4 lines (A, B, C and E) running in 10 minutes service during the day (between 05:30 and 19:00 hours) and 20 minutes service during the rest of the day. 2 other lines (G and H) run in 20 minutes service, and 1 line (F) run in 5 minutes service in the daytime and 10 minutes service the rest of the day. Some of the departures on line C are shortened. The route network is seen in figure 3.



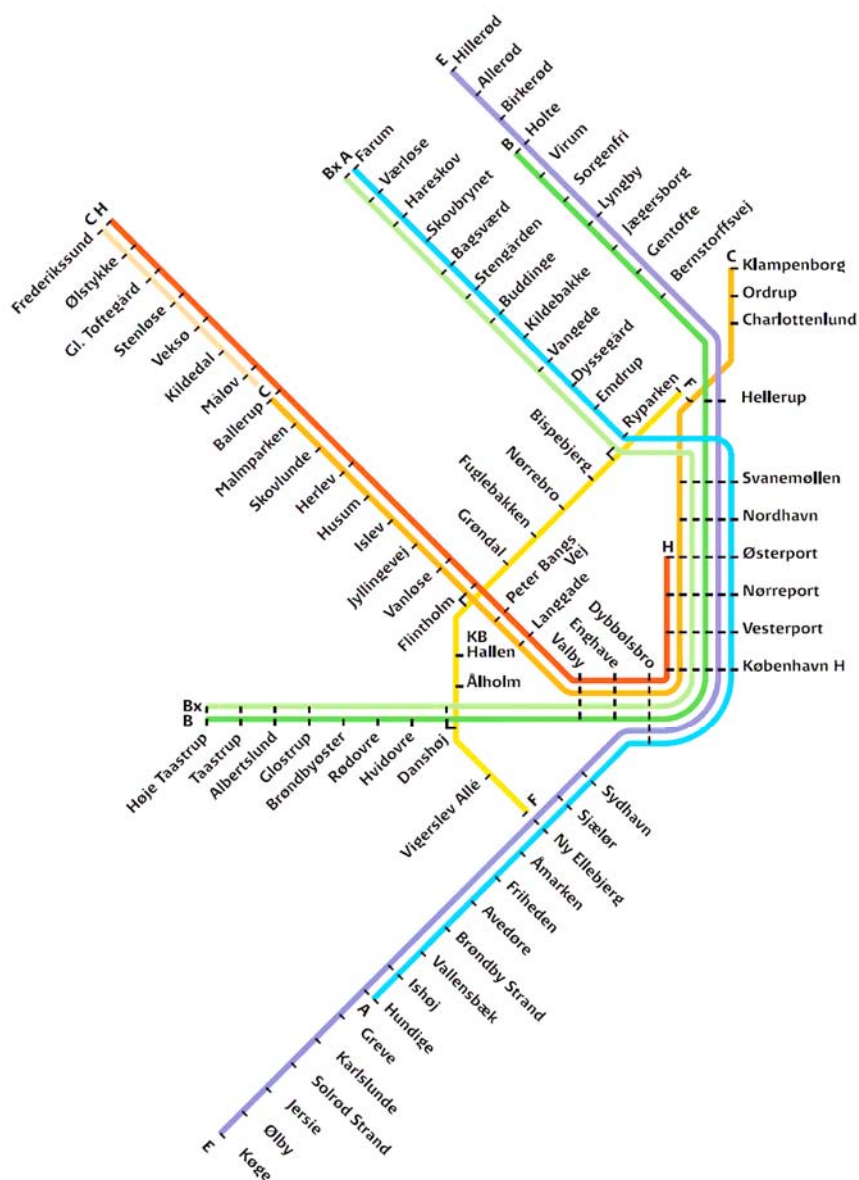


Figure 3: The Copenhagen suburban network, fall 2007.

5 Results of simulating disturbances on large a scale network

The RailSys model was run with 110 simulations, of which 2 contained deadlocks where trains blocked the way for each other. The remaining 108 simulations were used for further calculations and evaluations.



The results show that the regularity of the trains is higher than the regularity of the passengers, cf. figure 4. The traditional way of calculating passenger regularity (multiplying the delay of the train and the expected getting off the train) is demonstrated to result in higher passenger regularity than when calculated by the passenger regularity model. The differences between the regularity between trains and passengers are due to different numbers of passengers in the trains through the day. Furthermore, some passengers have to change from one train to another under the risk of missing the other train.

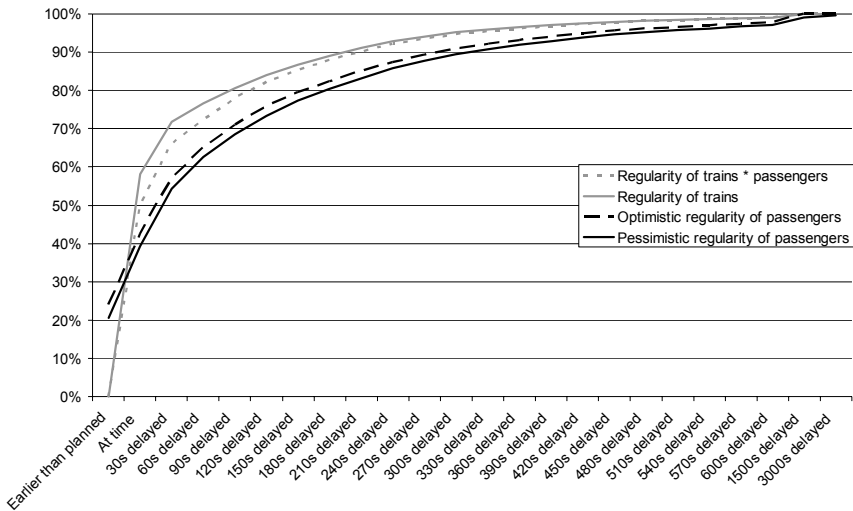


Figure 4: Regularity of trains and passengers at all stations.

Train delays do not necessarily cause passenger delays. Some passengers may even benefit from train delays. If a passenger arrives late to the station, a train delay may allow the passenger to catch an earlier train than expected. If the train catches up the delay, the passengers in the train may arrive on time. A similar situation may occur when a passenger changes from one line to another. If the train on the other line is delayed, it is possible to catch an earlier train than planned, thereby reducing the total travel time. In fact many passengers arrive earlier than planned (20 to 25 %), cf. figure 4 and figure 5.

From figure 4 it is seen that the optimistic regularity of passengers in general is higher than the pessimistic regularity of passengers. The difference can be explained by the passengers' knowledge of the delays. In the optimistic calculations, full knowledge of the delays in the entire rail network is assumed – to the extent that passengers have the information before the actual occurrence of the delays. In the pessimistic calculations passengers are assumed to follow a desired optimal route according to the timetable and only reconsider their route after a certain delay. Both principles of calculations have a certain error since passengers do not have full knowledge and passengers for some journeys choose the first train in their direction without waiting before reconsidering their route.



Thus, the true regularity of passengers is between the optimistic and the pessimistic values.

The distribution of arrivals at stations according to the planned journey (cf. figure 5 (a)) once again shows that some passengers arrive before scheduled (negative delays). However, it is difficult to see a difference between the result of the optimistic and pessimistic calculation of the passenger delays. This difference is seen in figure 5 b, illustrating a lesser tendency to delay and more passengers to arrive ahead of schedule when evaluated by the optimistic method rather than by the pessimistic method.

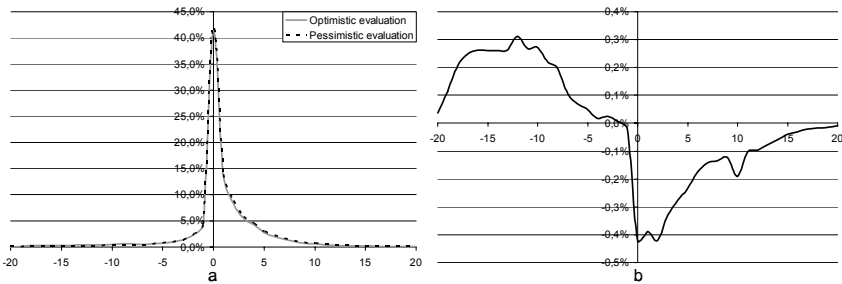


Figure 5: Distribution of arrivals according to the planned journey at all stations (a) and difference between optimistic and pessimistic evaluation of passenger regularity (b).

6 Discussion

Today, the passenger regularity model is run each night to evaluate passenger delays in the Copenhagen suburban network during the previous day [5]. The model has shown that passenger delays are larger than train delays [4], in accordance with the results presented in this paper. Other results (not published) show great similarity between the daily evaluation of delays and the simulated passenger delays.

Even though the RailSys model reproduces the results in Copenhagen quite well, the results can be improved. To do this and to improve reproducibility of the results, the RailSys model must be further calibrated to make the resulting delay at all stations similar to the daily operation. The RailSys model used in this paper has only been calibrated on an overall level so that the average delay for all stations is equal to the daily operation. It is very time-consuming, approaching the impossible, to gain exactly the same delay distribution as for the daily operation and the calibration should thus “only” be at the same level as (and not exact) the regularity of the daily operation [1].

When the RailSys model is calibrated, it is possible to evaluate the regularity of both trains and passengers at isolated stations as shown in figure 6. Beyond that, the passenger regularity model can be used for evaluating (and ranking) infrastructure improvements. The benefits for the passengers in terms of travel time and delays can be estimated and compared with the construction costs in,



e.g., a cost-benefit analysis. Furthermore, different candidate timetables can be evaluated and compared in the process of developing the best possible timetable for the passengers.

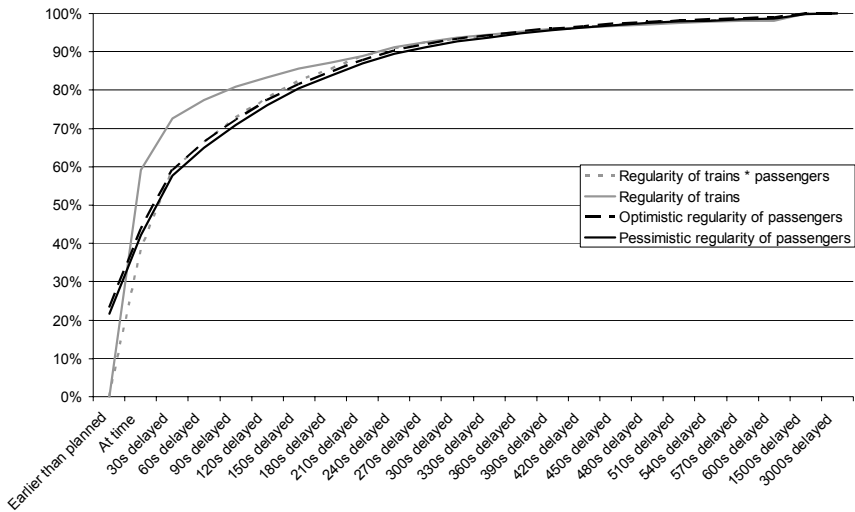


Figure 6: Regularity of trains and passengers at Copenhagen central station (København H).

Apart from the RailSys model, the passenger regularity model itself can be improved. According to the passenger regularity model, passengers will not change their route of travelling until a certain threshold of delay has been reached. However, on some stations or OD-relations, passengers will just take the first train in their direction. This phenomenon is characteristic for short journeys with high train frequency and is observed in the central Copenhagen between Østerport and Vesterport (cf. figure 3) with a train frequency of 2 minutes in each direction. The phenomenon might, however, also be observed at OD-relations with a lower frequency – i.e. Lyngby-Nørreport (cf. figure 3). Further work is necessary to estimate the correct threshold of delay to make passengers reconsider their route.

7 Conclusions

We have shown that it is possible to calculate the expected passenger delays by simulation of large scale networks and that there is a significant difference between train regularity and passenger delays.

The difference between the train regularity and passenger delays is due to the different number of passengers in the trains during the day and the fact that the passengers (to some extent) will change routes due to delays. Furthermore, there is a higher risk of delays in the rush hours due to more passengers and more trains.



The evaluation of passenger obtained with a simulation software, RailSys, and the passenger regularity model is comparable to the daily operation of the Copenhagen suburban network. Using a well calibrated RailSys model it will be possible to compare travel times and delays for different future scenarios – for changes in infrastructure as well as in timetables. In this way it will be possible to choose the best possible scenario.

Even though the results in this paper are very similar to what has been observed on the Copenhagen suburban rail network, the results can be improved both by better calibration of the RailSys model and estimation of the correct threshold of delay before reconsidering the route.

Acknowledgements

Rapidis Aps is thanked for the programming the passenger delay model. Rail Net Denmark (Banedanmark) is thanked for providing the infrastructure data for the Copenhagen urban rail network and discussions on the RailSys model. DSB S-tog (the train company of the Copenhagen suburban rail network) is thanked for providing the future timetables and OD-matrices for the travel patterns. Stephen Hansen and Kenneth Christensen, Centre for Traffic and Transport at the Technical University of Denmark, is thanked for developing the import-export tool from RailSys to the passenger delay model and the evaluation tool used.

References

- [1] Kaas, A. H., Punctuality model for railways. *Proc. of the 7th International Conference on Computers in Railways*, eds. J. Allan, R. J. Hill, C. A. Brebbia, G. Sciutto & S. Sone, pp. 853-860, 2000
- [2] Nielsen, O. A., A large scale stochastic multi-class schedule-based transit model with random coefficients. *Schedule-Based Dynamic Transit Modelling – Theory and Applications*. In *Schedule-Based Dynamic Transit Modelling: theory and applications*, eds. Wilson, N. and Nuzzolo, A. Kluwer Academic. pp. 51-77, 2004
- [3] Nielsen, O. A. & Frederiksen, R. D., Optimisation of timetable-based, stochastic transit assignment models based on MSA. *Paper accepted for Annals of Operations Research special issue on Optimisation in Transportation*. Forthcoming, Elsevier, 2006
- [4] Nielsen, O. A. & Frederiksen, R. D., Modelling train passenger delays. *Symposium on The Reliability of Travelling and the Robustness of Transport Systems*, eds. van Zuylen, H.J., pp. 137-156, 2005
- [5] Seest, E., Nielsen, O. A. & Frederiksen, R. D., Calculating passenger regularity in the Copenhagen suburban network. *Proc. of Traffickedays*, 2005 (in Danish).



This page intentionally left blank

Running time re-optimization during real-time timetable perturbations

A. D'Ariano¹ & T. Albrecht^{1,2}

¹*Faculty of Civil Engineering and Geosciences,
Delft University of Technology, The Netherlands*

²*"F. List" Faculty of Traffic and Transportation Sciences,
Dresden University of Technology, Germany*

Abstract

In the Netherlands the railway network is heavily used by heterogeneous train traffic and characterised by short headway times. That is why even small initial delays may perturb the timetable causing consecutive delays. In such conflict situations, traffic controllers have the complicated task to decide upon the optimal train schedule in real-time. They could be assisted by sophisticated conflict-solving systems. Optimal train running profiles can be designed which fit better to the new train order and allow reducing delays and energy consumption at the same time. The paper presents a formulation for this complex problem that makes it suitable for quantitative analysis and for optimizing actual running times at network scale. A conflict solution system is developed, that models the train scheduling problem as an alternative graph. Adopting the blocking time model, safe headway distances between trains are assured by any real-time traffic control measure. The optimal solution from a network point of view can be improved by modifying the speed profiles locally for the individual train routes. A constructive heuristic algorithm for the dynamic modification of running times during operations is proposed that satisfies the timetable constraints of train orders and routes and guarantees the feasibility of the running profile, while taking into account the properties of the signalling and train protection systems in use. A real-world example from the Netherlands in case of the Dutch signalling system NS54 is presented to demonstrate the benefits of the proposed methodology.

Keywords: railway traffic management, train scheduling, delay minimization, train speed optimization, energy optimal train control.



1 Introduction

The off-line development of detailed conflict-free timetables is a complex and recurrent problem, and typically requires many months. Since disruptive events are unpredictable time reserves are distributed over the timetable, preferably guided by past experience and empirical data [1]. During operations, however, unforeseen events may disrupt the timetable and cause conflicts between train paths which must be resolved in real-time. Moreover, the railway infrastructure is becoming more and more saturated by which local delays are more difficult to manage and easily generate many consecutive delays (due to conflicts).

Current operational traffic management is reactive: A train driver tries to adhere to the original schedule and responds to the signalling system when the route ahead happens to be occupied. Dispatchers only reschedule the route setting plan when trains have a considerable delay, and (network) traffic controllers become active when train traffic already is highly disrupted. Hence, improving reliability of dense train traffic requires an advanced railway traffic management system that accurately monitors current train positions, predicts potential conflicts and reschedules trains in real-time such that consecutive delays are minimized [2].

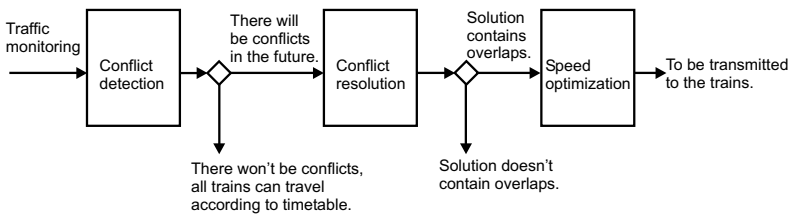


Figure 1: Components of a proactive traffic management system.

The introduction of computerized on-line decision support systems aims to prevent the decision maker from taking wrong decisions. Such systems should be designed to support operators to quickly re-schedule trains during real-time perturbations and typically contain the following components (cf. Fig. 1):

1. Traffic monitoring: Take as input the position, the speed and the planned timetable for each running train in the railway network.
2. Conflict detection: Given the current infrastructure status, actual timetable and current train delays and speeds, find potential conflicting train paths that require the same infrastructure element at the same time (*conflicts*): this will be prohibited by the safety system.
3. Conflict resolution: Given the current train delays and predicted conflicts, find an optimal solution by rescheduling and/or rerouting trains. This solution must respect the constraints of the signalling system in use: A train that gets too close behind another is forced to slow down by the signalling/train protection system in use. Getting into such a situation corresponds to an *overlap of blocking times* according to blocking time theory [3].



4. Speed optimization: Send new targets (time and advisory speed at key locations) to rescheduled trains with the aim of respecting punctuality and saving energy.

Many consecutive delays can be prevented if traffic is proactively managed. Based on an accurate monitoring of actual train positions potential conflicting routes can be predicted in advance and resolved in real-time. The adjusted targets (location-time-speed) have then to be communicated to the relevant trains. Driver Assistance Systems [4] installed onboard can support the driver control its train best to reach the overall optimum. Such systems have to consider the behaviour of the safety system to avoid undesired safety braking, consequent slow downs and higher energy consumption of trains. They will allow not only more precise control of trains, but also effective train speed coordination on open tracks, securing time windows at junctions/crossings, or synchronizing arriving trains at stations in case of delays and expected route conflicts.

The literature on traffic management focuses on conflict detection and resolution (CDR) systems (cf. [5] for an overview). Only few studies are known which take into account the possibility of speed coordination to improve the rescheduling solution: Asuka and Komaya [6] mainly focus on urban rapid transit with low speeds and short distances between stations. Huerlimann [7] focuses on speed optimization at a single junction. This paper presents an advanced traffic management system which is able to consider a complicated network for the CDR system and improve the obtained solution by speed optimization in network nodes afterwards.

The paper is organized as follows: Section 2 introduces the CDR system. Section 3 deals with the optimal speed control in conflict situations. In Section 4 we describe a test case based on the Utrecht-Geldermalsen railway link. Finally, conclusions are drawn and future research is discussed.

2 Conflict detection and resolution system

The aim of real-time rescheduling is to control the railway traffic in case of perturbations. The CDR system used here models the railway system as an alternative graph [8]. This graph represents the train paths of all trains in a given control area along with their precedence constraints (minimum headways computed using the blocking time model [3]). Moreover, special (alternative) arcs represent operational choices such as the train order at a crossing or merging section. A decision is made by selecting one of two alternative arcs which then fixes a precedence constraint between two trains at a potential conflict point. This alternative graph can be used to model different signalling systems and therefore is quite flexible [9]. In case of fixed block signalling each block signal corresponds to a node in the alternative graph and the arcs between nodes represent blocking times or headway times. A feasible schedule assigns passage times to each node such that all precedence constraints are satisfied. The CDR problem (or train scheduling problem) can then be defined as follows: For given initial delays find the shortest paths in the alternative graph and fix the corresponding selected alternative arcs.



All trains are considered simultaneously with the aim to minimize the maximum consecutive delay in the network. D'Ariano et al. [10] developed an efficient branch and bound algorithm to find an optimal schedule using several speed ups exploiting the problem knowledge.

The alternative graph model assumes deterministic blocking and waiting times and thus does not take deceleration and acceleration into account in case of hindrance. D'Ariano and Pranzo [11] therefore developed a train management system that updates the speed profiles of trains according to the signal aspects. In each step of an iterative rescheduling procedure the CDR problem is solved and subsequently it is checked whether train paths overlap (yellow signals). If overlaps arise the speed profile of the second train (that passes a yellow signal) is updated and the blocking times are changed accordingly. In the next iteration possible new 'queueing' conflicts are solved, which again may lead to slowing down of trains. The rescheduling procedure terminates in a finite number of iterations and gives a conflict-free timetable with admissible train dynamics.

3 Optimal speed control in conflict situations

The solution of the CDR system may contain overlaps. That means in practice, that the concerned train will have to reduce its speed involuntarily. This may cause operational disadvantages:

1. Traditional ATP systems (like the Dutch ATB-EG) will force the train driver to decrease the speed of its train in any case in order to be able to reach a safe state before the next signal, which is expected to be the limit of movement authority but whose position is unknown to the onboard unit of the ATP system.
2. Train speed can only be increased again, if a signal upgrade has been recognised by the ATP system. Whereas modern ATP systems like ETCS level 2 can do so automatically in a safe manner, old systems rely on a non-safe confirmation of the signal upgrade by the driver. At the time the conflict is over, the train will have a lower speed than originally planned in the timetable. In order to get back to its travelling speed it has to re-accelerate, which costs time in the first place (and therefore causes additional delay) and energy in the second place.

In order to reduce those negative effects, a constructive heuristic algorithm is proposed for the computation of the optimal train trajectories. This algorithm is illustrated in Fig. 2. It is based on the idea that the train is slowed down slightly some time before the possible overlap, in order to be able to re-accelerate without being hindered and reach the conflict area with optimal distance behind the train causing the conflict. The train should then never have to pass a signal aspect that forces a modification of its optimal trajectory.

Of course, the algorithm only has to be started in case that overlaps (speed decreases due to signal aspects) exist; assuming the train goes from its state it enters the corridor on its planned (energy-optimal) path. If that is the case, step 1 of the algorithm (Fig. 2) consists of the computation of the fastest possible trajectory



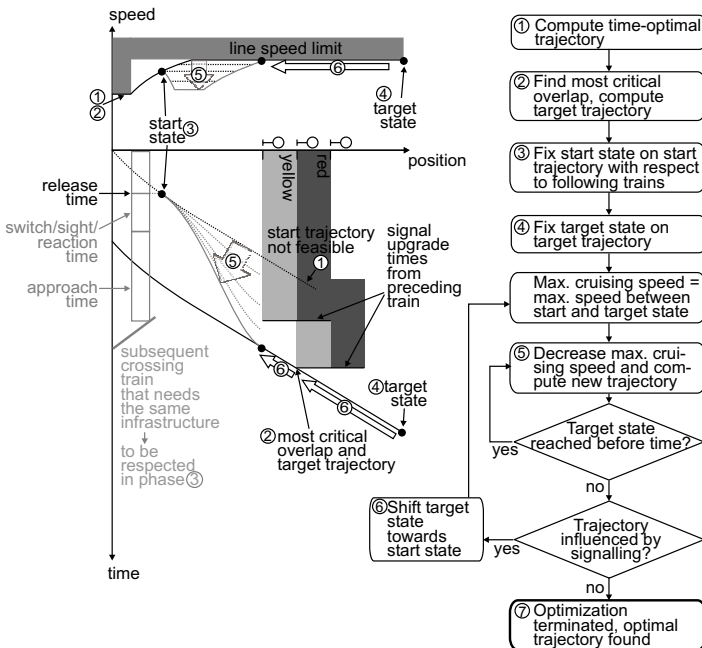


Figure 2: Flow diagram and the algorithm steps illustrated in a theoretical example.

from the start state of the train (position, speed) to its target state (scheduled passing time) at the end of the corridor without regard of the signalling system. For this trajectory, the most critical conflict can be determined (biggest negative difference between passage time of the time-optimal trajectory and signal upgrade to undisturbed operation - green signal aspect). Step 2 consists of determining the so-called target trajectory by shifting the time-optimal trajectory in time in such a way, that it passes the critical conflict at the time of signal upgrade. It shall now be assumed that the arrival time of this trajectory is later than the planned arrival time, so this target trajectory is the only solution with minimal delay which does not have to pass a yellow signal (there may be other solutions with smaller delay at the end of the corridor, which necessitate passing a yellow signal once [12]).

The next steps of the algorithm consist of finding the best transition between the expected trajectory of the train (1) and the target trajectory (2). It shall be computed according to the following criteria:

- The train shall be slowed down the least possible, because every slowing down needs to be compensated by re-acceleration which costs energy.
- The smaller the distance and the longer the time to the critical conflict, the more the train has to be slowed down.
- The earlier a train diverges from its original trajectory, the later it will be at any given position before the conflict. Those delays may cause consecutive delays of subsequent trains.



Those three factors have to be taken into account in the construction phase of the transition trajectory. In order not to hinder subsequent trains which partially use the same infrastructure, step 3 sets a start state on the original trajectory. The time the train may diverge from its trajectory can be computed using blocking time theory recursively (from the time the next train will need to use the infrastructure back to the required release time of the same infrastructure for the examined train).

In step 4, the target state is fixed on the last point on the target trajectory, which theoretically would allow having the maximal distance between start and target state. Then, the transition trajectory must be computed (step 5). It must contain minimum three regimes and two switching points [12]. Here, three phases are proposed: A regime of braking/ acceleration, cruising at a given cruising speed and re-acceleration to the target speed. To find the switching points between the regimes, the maximal cruising speed between start and target state is successively reduced, the first and last phase are prolonged respectively. The arrival time at the target state increases and step 5 is repeated until the target state is reached on time.

It must then be checked, whether the train can follow this transition trajectory with respect to the signalling system (step 7) or whether it had to brake involuntarily when following it. In the latter case, the target state is moved towards the start state (step 6). That generally leads to lower cruising speeds of the transition trajectory, where the risk of hurting the signalling constraints decreases. If the target state is moved too much towards the start state, it may be possible that no feasible solution for step 5 exists and the desired target trajectory can not be reached under the given constraints. Either the constraints have to be redefined (move start state towards section entrance, step 3) or the target trajectory moved forward in time and the optimization has to be re-started. In case the computed target trajectory arrives too early at the exit of the regarded corridor the available running-time reserve must be distributed either before or after the conflict. That may be done using Dynamic Programming [4].

4 Case study

This section illustrates the influence of the effective railway traffic management system proposed in the previous sections by means of a real-word example. A node of the existing Dutch railway network between Utrecht and Geldermalsen (Fig. 3) is regarded. It is composed of a main corridor of around 14 km and includes Culemborg station. There are four trains in the network. Train A is a freight train, going from Culemborg (block section 15) to Utrecht (block section 22). Train B and C are intercity passenger trains going from Geldermalsen (block section 1) to Utrecht. Finally, train D is a regional passenger train going from Geldermalsen to Dordrecht (block section 8). Train A shall be delayed and the running of the following trains on the corridor is influenced by the presence of permissive signals and signalling block constraints at the entrance of Culemborg station. Potential conflicts can also be found at the exit of Culemborg station. Trains A, B and C share block sections 15-22. Since trains B and C follow the same path they share all the block sections. Train D shares block sections 5 and 6 with trains



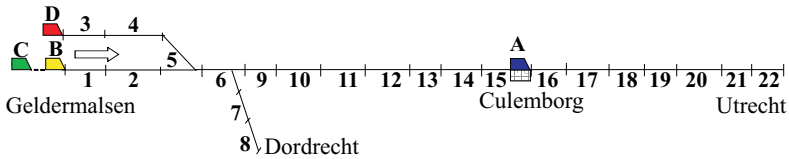


Figure 3: Utrecht–Geldermalsen railway link.

B and C. Note that due to the initial position of trains, train C is not allowed to precede trains B and train A can not be surpassed, and therefore we have a priori order among those trains. The switching time to release or set the interlocking routes and signals is taken to be one second, on the basis of automatic signal blocks with electronic technology. Furthermore, sight and reaction times at sight distance of the approach signal are assumed equal to 14 seconds. The track speed limits are 80 km/h for the sections 1-8, 130 km/h otherwise. The signalling system consists of two speed levels in case of a conflict ahead (80 km/h, 40 km/h), which are signaled depending on the section lengths. The Dutch ATP-System ATB-EG requires immediate braking after entering a section, where the signaled speed is lower than track maximal speed, which is also taken into account in the simulation.

4.1 Conflict detection and resolution system solution

The solution obtained by the branch and bound algorithm is given in Fig. 4. All train paths involved in the example are represented in terms of blocking times. Due to the large input delay of train A, re-timing decisions have been taken to obtain a feasible schedule with respect to headway and signalling constraints but the system solution presents the train ordering of the original timetable. In detail, trains B and C interfere on block sections 15-17 as illustrated by the overlapping blocks. In this situation, the planned speed profiles of those trains can no longer be used in the CDR algorithm. They are adapted according to the iterative rescheduling procedure of D'Ariano and Pranzo [11].

In this case, delayed trains run at their maximum allowed speeds and in case of a yellow (or red) signal aspect an ordinary driver behaviour is applied: At sight distance of the approach signal drivers are supposed to start decreasing train speed. Train D is not delayed in this example thus maintains its original speed profile.

4.2 Speed optimization system solutions

For the example of Fig. 3, three different solutions in terms of train speed trajectories shall be discussed. The solution given by the CDR system and two variants of speed optimization are proposed in Fig. 5. In all three cases, train A runs at maximum speed, trains B, C and D adopt different speed profiles.

The CDR solution presents feasible speed profiles for each train (Fig. 4). At the exit of the network train B (C) has a delay of 128 (167) seconds and train D is



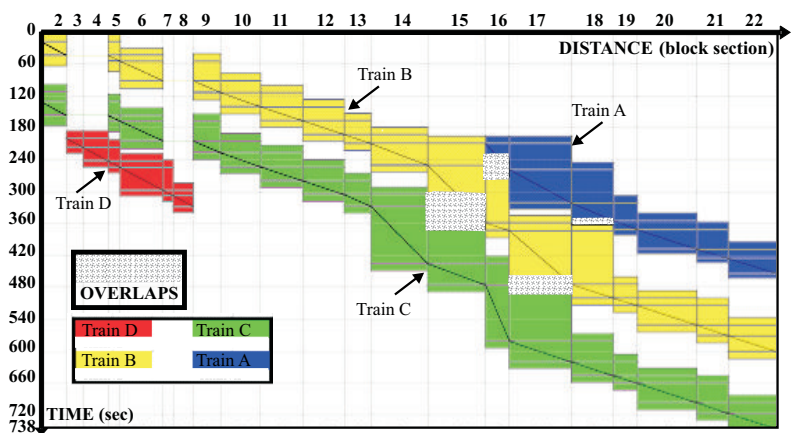


Figure 4: Blocking time plot of the CDR solution.

not delayed, in case the drivers are supposed to react to the current state of the signalling system only.

The next solution represents a speed optimization which only takes into account the previous train and not subsequent trains. Here, step 3 of the algorithm in section 3 is not considered and trains B, C and D are optimized in the order that they pass the critical section. With the aim of minimizing energy consumption of the individual trains, trains B and C are slowed down as soon as they enter the corridor. As a consequence train D can not get out of the situation but delayed. In detail, with respect to the previous solution the energy consumption for all trains is reduced by 45%, whereas the sum of delays is about the same (reduced delay for trains B and C and increased for train D). For train D, this represents a deterioration of the solution quality compared to the CDR solution without speed optimization.

A second speed optimization run is done, this time respecting the constraints given by train D when fixing the starting states of train C and B in step 3. The possibility of speed decrease of train C is limited to the sections, where it doesn't influence train D. Train B must be re-optimized in order not to disturb train C. With this solution, the regarded trains B, C and D need 30% less energy than without speed optimization. Train D is not disturbed and therefore, overall delay can be reduced by 45 seconds.

The two proposed speed optimization solutions are pareto-optimal: Min. energy consumption and min. delay respectively. Traffic operators must discuss which of the solutions fits their objectives during operations.

5 Outlook and conclusions

This paper showed the effectiveness of an intelligent method to reduce consecutive delays by identifying potential conflicts, rescheduling train traffic and optimizing the speed trajectory of each train involved in a conflict/overlap situation. Such



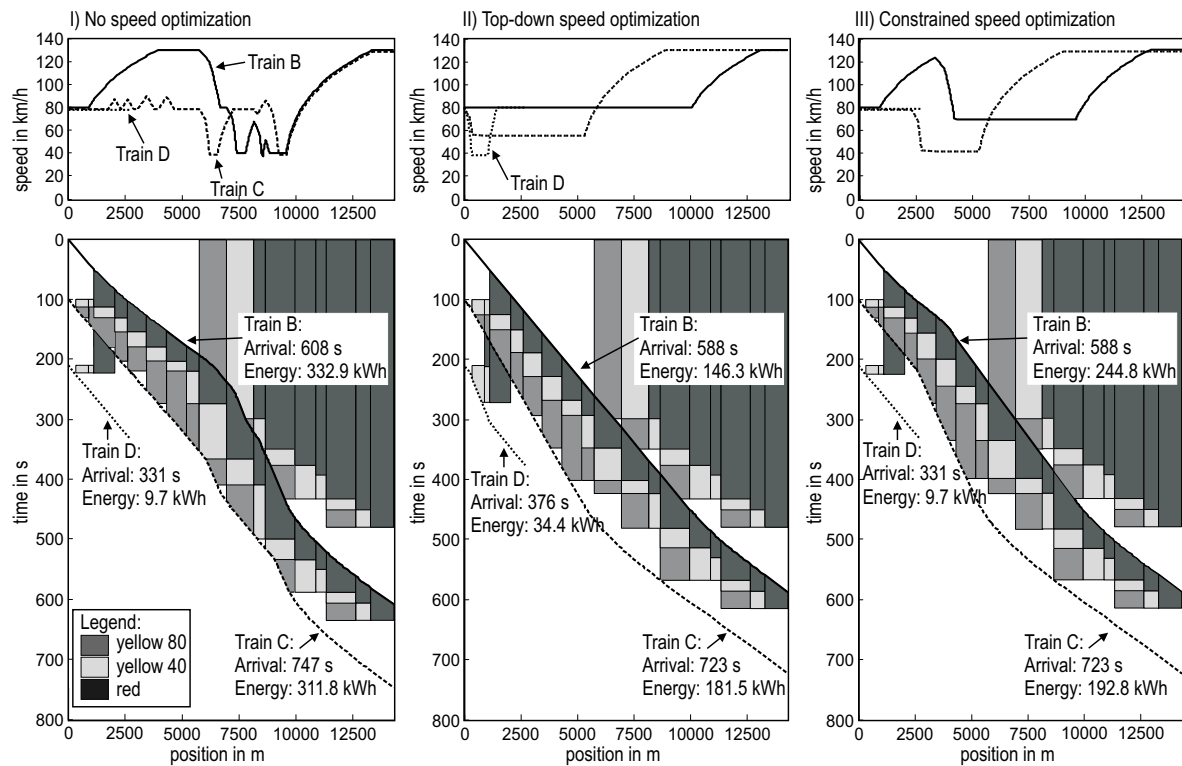


Figure 5: Comparison of three variants of train speed trajectories.

a system can be implemented in an advanced traffic management system which improves operations reliability with small investments compared to the alternative of building new infrastructure.

A case study of the Utrecht-Geldermalsen railway link showed how the CDR system and the speed optimization can be adopted in order to react proactively when signal aspects change. The obtained solutions clearly indicate the need to develop on board driver support systems in order to improve punctuality and minimize energy consumption.

In the future, it is planned to include the speed optimization into the iteration algorithm of the CDR system already to be able to better estimate the consequences of optimal speed control at network level.

References

- [1] Hansen, I.A., Increase of capacity through optimized timetabling. In: *Computers in Railways IX*, pp. 529-538, Southampton, 2004. WIT Press.
- [2] Ho, T.K., Norton, J.P. & Goodman, C.J., Optimal Traffic Control at Railway Junctions. *IEE-Proc.-Electr. Power Appl.*, **144**(2), pp. 140-148, 1997.
- [3] Pacht, J., *Railway Operation and Control*. VTD Rail Publishing, Mountlake Terrace, USA, 2002.
- [4] Albrecht, T. & Oettich, S., A new integrated approach to dynamic schedule synchronisation and energy-saving train control, In: *Computers in Railways VIII*, pp. 847-856, Southampton, 2002. WIT Press.
- [5] Oh, S.M., Hong S.H. & Choi, I.C., Railway conflict detection and resolution in the Korean railway system. In: *Computers in Railways IX*, pp. 675-684, Southampton, 2004. WIT Press.
- [6] Asuka, M. & Komaya, K., Automatic train operation using braking pattern transitional time. In: *World Conf. on Railway Research*, pp. 628-636, 2003.
- [7] Huerlimann, D., *Objektorientierte Modellierung von Infrastrukturelementen und Betriebsvorgängen im Eisenbahnwesen*. PhD Thesis, ETH Zurich, 2001.
- [8] Mascis, A. & Pacciarelli, D., Job shop scheduling with blocking and no-wait constraints. *European Journal of Operational Research*, **143**(3), pp. 498-517, 2002.
- [9] Giannettoni, M. & Savio, S., The European Project COMBINE 2 to Improve Knowledge on Future Rail Traffic Management Systems. In: *Computers in Railways IX*, pp. 695-704, Southampton, 2004. WIT Press.
- [10] D'Ariano, A., Pacciarelli, A. & Pranzo, M., A branch and bound algorithm for scheduling trains on a railway network. *Technical Report DIA-97-2005*. Dipartimento di Informatica e Automazione, Università Roma Tre, 2005.
- [11] D'Ariano, A. & Pranzo, M., Conflict resolution and train speed co-ordination for solving timetable perturbations. In: *1st Int. Seminar on Railway Operations Modelling and Analysis*, 8-10 June 2005 Delft, The Netherlands.
- [12] Albrecht, T., Energiesparende und verspätungsminimierende Zugsteuerung in Konfliktsituationen. *Signal und Draht*, **97**(12), pp. 18-22, 2005.



ALFa – a software tool for optimal scheduling of demand oriented train services

S. Scholz & T. Albrecht

Dresden University of Technology, Germany

Abstract

A scheduling strategy considering demand variation during the day and within the network is outlined. The optimal headway is calculated taking into account demand and supply simultaneously achieving equilibrium. The train service structure is complicated due to different headways for each hour of the day and various parts of the network. Short train headways shall be provided for sections with high demand whereas sections with lower demand (outskirts) request easily recallable departure times. That is why the use of Genetic Algorithms is proposed to transform such an operating scheme into an optimal train service. They allow to find a timetable with non-conflicting train paths, which is characterized by even headways for all sections of the network. The timetable can also be optimized for minimal fleet size or both goals simultaneously. In order to ease the application of such operating strategies, the proposed algorithms were implemented into a computer-aided planning system called ALFa which is described in this paper.

1 Introduction

Demand-driven train operation of fully automated underground railway lines can provide capacity strictly according to the fluctuation of demand. This helps avoid overcapacity during off-peak times and improves the revenue-cost-relation of the operator due to reduced operational efforts. Since this strategy has only been applied to single lines yet [1] the feasibility must be proven for a more extensive network of lines and their temporal and spatial demand distribution.

The basic approach of demand-driven operation is characterised by the control of the train headway T_S in such a way that transport supply (capacity) is adapted to demand as much as possible. In the early morning and the late evening the longest train headway is chosen short enough so that the passengers do not need to



know the timetable. Usually, a headway between six and ten minutes is used (cf. system VAL in Lille and Rennes or Skytrain in Vancouver). During the peak hours and during the day the train service is adapted to demand by varying the headway between one and six minutes (cf. upper left subfigure of Fig. 1).

Apart from this strategy to adapt to the temporal demand variation several operators of automated rapid transit systems also make use of a spatial adaptation. Short-turn trips serve only parts of the whole network and some sections of the network are commonly served by multiple lines. Furthermore, the vehicle size can be adopted in the course of the day or within the network (cf. Fig. 1). In practical operation even a combination of these four techniques is used. This paper focuses on the combination of temporal and spatial adaptation of the train headway.

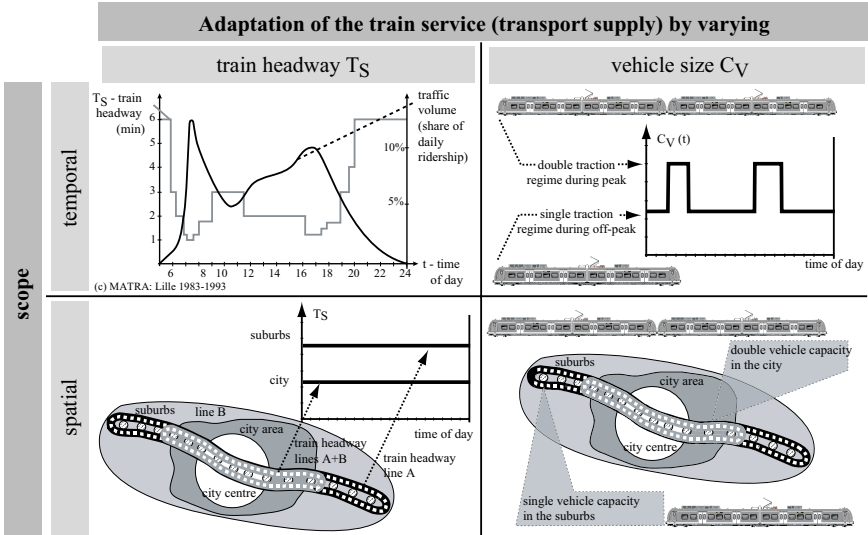


Figure 1: Adaptation of train service to demand - operational concepts.

Although a multitude of timetabling systems are described in the literature, none of them was able to solve the above problem of temporal and spatial adaptation of train headway on a network of lines in a satisfying, automatic manner. This paper presents a prototype of an integrated timetabling system, that has been developed at Dresden University of Technology and incorporates for the first time a supply-demand model and a timetabling algorithm in one tool. They will be described in section 2 and 3 of this paper. Section 4 and 5 present the developed software prototype and its first application.



2 Train service with supply-demand equilibrium

The adjustment of supply to spatio-temporal variations in passenger demand helps improve basically the so called traffic efficiency [2]

$$\eta = \frac{vl(T_S)}{bl(T_S, C_V, \gamma)} \cdot 100\% \triangleq \frac{[vl]}{[bl]} = \frac{\text{pass.-km}/(\text{year and track-km})}{\text{place-km}/(\text{year and track-km})}. \quad (1)$$

The operational effort bl is derived from the transport supply which is determined by the train headway and the vehicle size C_V (regarding a maximum load factor γ). It can be adjusted to demand by choosing the optimal train headway T_S . Since line-haul rapid transit systems are considered, this train headway must be chosen with respect to the heaviest loaded section $x = \mu$ of a line. The traffic performance vl (demand) is influenced by the quality of the train service, e.g. the train headway, too.

In order to achieve a high efficiency of the operation (cf. eq. 1) the transport supply and the transport demand must match optimally [2]. The maximum of the efficiency can be obtained by minimizing the deviation ε_{VAG} between supply and demand at the heaviest loaded section of a line.

$$\varepsilon_{VAG}(t, T_S, \mu) = \frac{C_V \cdot 60}{T_S} \cdot \gamma - VA(t, T_S, \mu) \rightarrow \text{MIN} \geq 0. \quad (2)$$

Since demand and supply interact with each other, this supply-demand response must be considered in order to achieve consistent planning results. An aggregated transport demand model taking into account the response of demand to changes in supply was developed and described in detail in [3]. With the help of this model optimal train headways can be calculated that minimise overcapacity.

The dependency between demand and supply is illustrated in Fig. 2 for the suburban railway line of a medium sized conurbation [3]. It approx. corresponds to the city of Dresden, Germany and its suburbs. The interception point between supply and demand curve indicates a train headway with minimum overcapacity.

This transport model and the derived headway control strategy deliver the optimal train headway for each hour of the day and each branch of the network. This nominal train service structure with optimum supply-demand adaptation needs to be transferred into a feasible timetable with detailed departure times of the trains for each station.

3 A Genetic Algorithm approach for automated timetabling

3.1 Problem description

An optimal departure time within the given hour must be found for each of the obtained train rides. This is done in two stages:

1. Trains on lines with low operating frequencies are scheduled to fixed and easily recallable departure times (fixed headway over several hours).



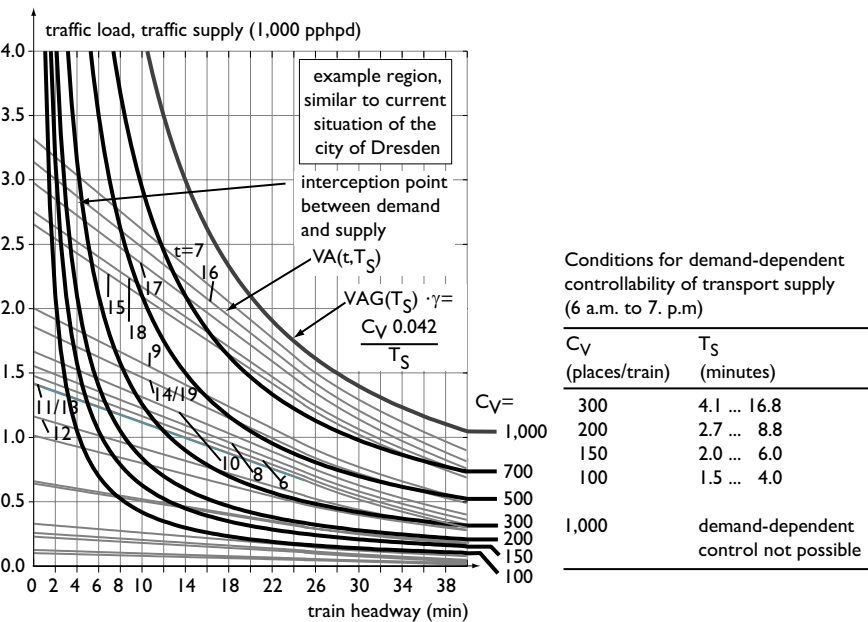


Figure 2: Conditions for demand-dependent controllability of transport supply [2].

2. Trains on all other lines, especially on those serving only small parts of the network, are scheduled in such a way, that train headway becomes as smooth as possible on every branch of the network and the necessary fleet size to realize this timetable becomes minimal.

The second task is very complicated, especially because of the long travel times on some of the lines. Those are the reason, that the global optimum of this problem can only be found by taking into consideration the whole operating day and not the single hours individually.

During the timetable design process, the most important constraint to be respected is minimal headway. It must be assured between suburban trains following each other as well as between suburban trains and trains of other operators running on the same infrastructure (freight or long distance passenger traffic).

3.2 Computing optimal departure times

Taking into account the nature of the cost function as well as the important number of variables, Genetic Algorithms (GA) are proposed to search for the optimal solution: This method has been described in detail in the literature [4, 5] and has become popular to solve complex railway optimization problems (e.g. [6, 7]). The fundamental way of solving technical problems by imitating the process of natural evolution is illustrated for the timetabling problem in Fig. 3.

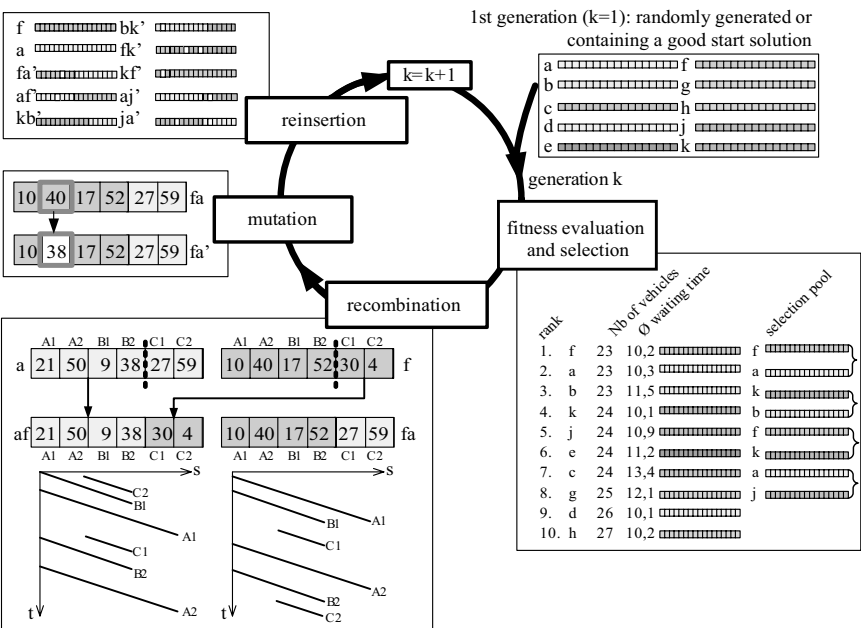


Figure 3: Using genetic algorithms for timetabling.

The most challenging point when using GA is to find a suitable coding of the problem to be solved. In Fig. 3, one way is described: the departure time (given in full minutes or part of minutes within the hour of the train's ride) is coded directly on a chromosome. When using this coding with the randomly acting GA operators, it is likely to happen that the timetable read from the chromosome contains departure times which do not respect the minimal headway. Those invalid timetables can either be penalized or transformed into feasible ones, which is proposed here.

Because of this problem of invalid solutions, an alternative coding could consist of taking the train headways as independent variables as well as the order of the lines. This coding reduces the number of conflicts drastically, but not to zero, because of the fixed departure times of trains of other operators that have to be respected. Both codings offer different advantages as was shown in [8].

The transformation process results in a timetable without conflicts. The cost function (passenger waiting time, fleet size) can now be calculated. Due to the nature of GA (which work on a representation of the solution and not directly on the parameters of the solution), other optimisation criteria (nb of drivers, connections to other lines) can be easily incorporated without changing the algorithm or the coding.

4 Case study and test scenarios

The outlined approach was implemented into a software prototype. The first application was the suburban railway network of Dresden, Germany. Here, the following operation scenarios have been tested (cf. Fig. 4):

- *status quo scenario*: situation of the year 2000 with network-wide headways of 20 to 30 minutes,
- *M1-ITC scenario*: traditional fixed interval timetable (peak: 7.5 min.; off-peak: 15 min.) and some short-turn trips during the peak period (track infrastructure: ITC - intermittent train control),
- *M2-ITC scenario*: advanced M1-ITC scenario with shortest train headways of 5 minutes and demand oriented train service in the city centre,
- *M3-ETCS scenario*: minimum train headways of 3.8 minutes, combination of temporal and spatial adaptation to demand in the urban area,
- *AGT scenario*: further enhanced M3 scenario assuming automated train operation with shortest headways of 2 minutes and network-wide combination of temporal and spatial adaptation to demand

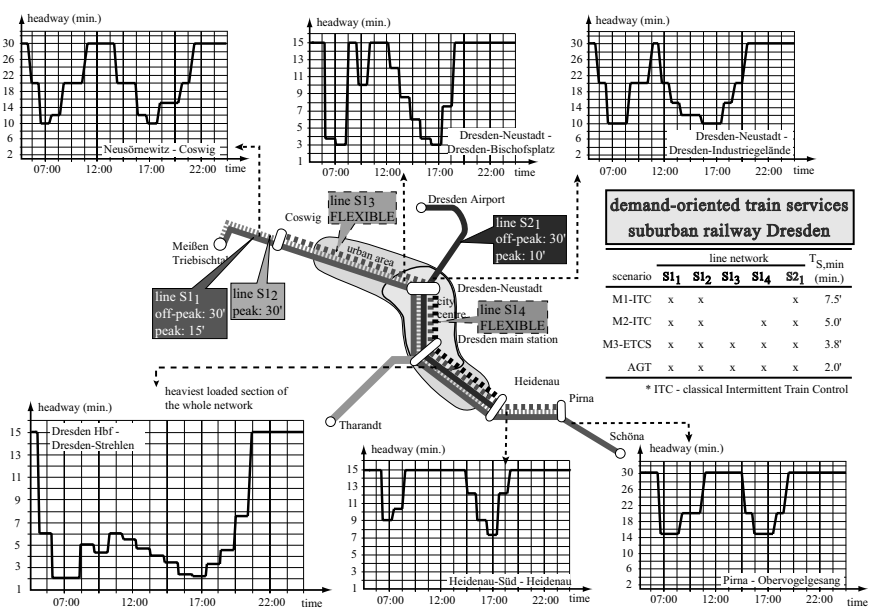


Figure 4: Train service structure: line network and headway for different scenarios.



5 Software prototype

5.1 Scope of application

The developed software solution aims to solve three planning tasks:

1. ALFa-E: Event-based (re)-scheduling of existing timetables. The modification of existing timetables is especially important for hardly predictable changes in demand (weekend trips due to nice weather, shift from car to public transport due to frozen roads etc.). External short term demand forecasts can be used to adapt the nominal long term train service. Minor modifications include some extra trains or a reduction of the number of trains if possible. It can also be used in case of technical problems, e.g. blocked tracks due to switch failure.
2. ALFa-M: Medium term planning. The planning of demand-oriented train services several months in advance is considered as medium term planning. The nominal timetable for the next operating period can be adapted to average changes in demand in the course of the day and changes on weekdays and on weekends. The scenarios *M1* and *M2* belong to this category.
3. ALFa-S: Strategic planning. The third objective of the ALFa tool is dedicated to general planning issues with strategic goals and basic improvements due to major modifications of the train service and even the rolling stock. This fundamental modification must be considered a long term strategy requiring adequate fundamental studies on its expected benefits and operational feasibility. The scenarios *M3* and *AGT* are examples of this sort of planning problems. Especially this issue requires to consider the supply-demand response of such basic improvements of the quality of train service. These fundamental modifications lead to highly complicated (time and space variant) train service structures which can hardly be transferred into feasible timetables with manual effort only.

5.2 Functionality and graphical user interface

The software prototype includes a transport demand module which enables the user to design the train services taking into account the demand reaction for that service' quality. Furthermore external demand data can be loaded through an interface (cf. Fig. 5). The train service is determined by the vehicle type of each line, the desired maximum load factor of the trains, the minimum and maximum admissible headways and the assumed supply-demand response. As a result the number of trains per hour for each section of the network is obtained.

Furthermore, all standard tasks for railway timetabling (cf. Fig. 6) such as: modeling of infrastructure, management of train running times and nominal station dwell times, support of various rolling stock, fixed timetables for certain lines (long-distance trains) and vehicle assignment are included.



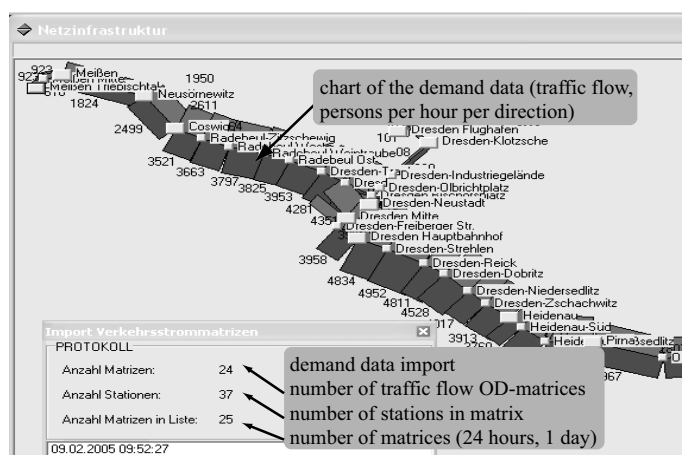


Figure 5: Network with demand distribution.

Apart from the implemented optimisation algorithms the software includes further functions that are primarily important for convenient user interaction. The calculated timetables can be displayed both as scalable time-distance-graph and spreadsheet (cf. Fig. 8) with export interface into standard EXCEL format.

The software manages all operational and infrastructural data which can be saved to different projects for easy modification and reuse.

The basic steps which need to be taken to calculate the timetable are as follows: definition of the track network with stations and interconnections, definition of a line network, definition of vehicle types, import of the transport demand (cf. Fig. 5) or explicit calculation of the required supply in order to meet demand, presetting of line parameters (planning constraints, cf. Fig. 7) and initiation of the automatic calculation of a timetable. An example timetable of the AGT scenario is displayed in Fig. 8.

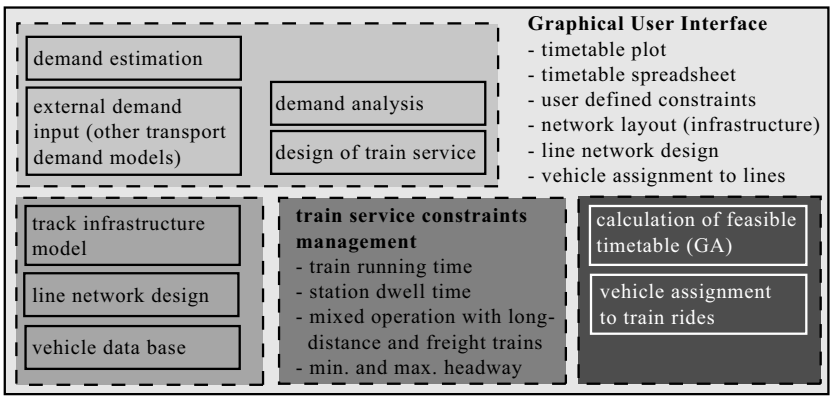


Figure 6: Modules of software prototype ALFa.

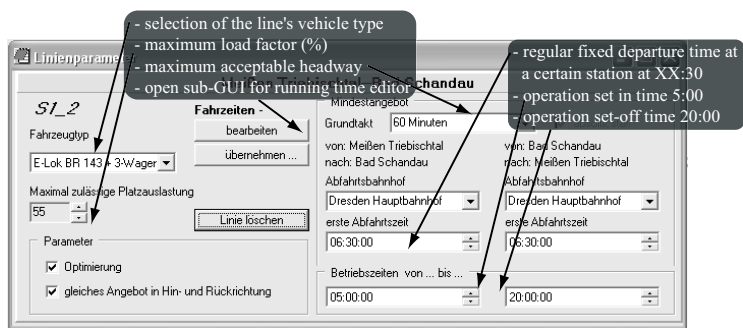


Figure 7: Basic line characteristics as calculation constraints.

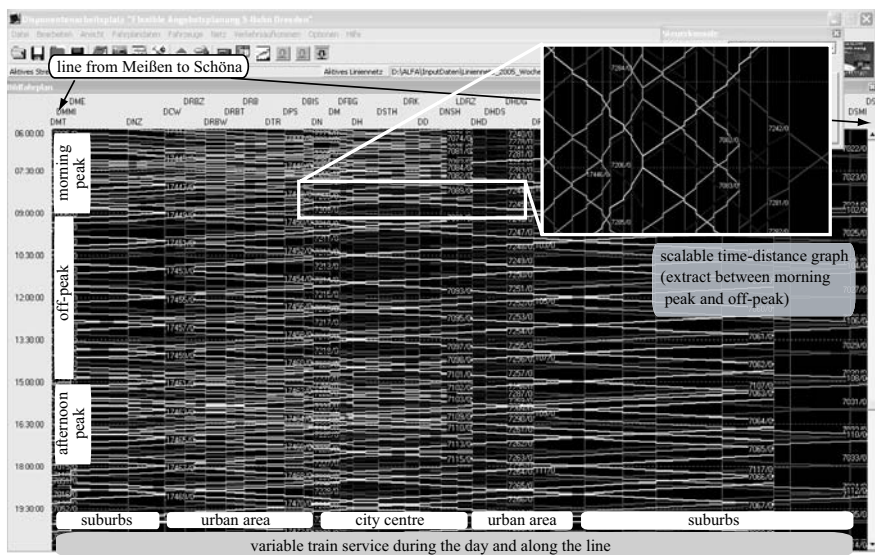


Figure 8: Time-distance train graph of the AGT scenario.

6 Conclusion

The proposed compact transport demand model makes it possible to merge transport demand modeling and train service planning for the first time into a single train timetabling system. Planning of new and basically changed train services becomes more realistic and can be considered superior to straight-forward planning approaches with fixed demand. Genetic Algorithms are suitable to solve the difficult timetabling problem with the different operational constraints. The developed software prototype demonstrated its usefulness in several case studies for the suburban rapid rail system of a medium sized conurbation.



Acknowledgement

ALFa was developed for DB Systems GmbH/ DB Regio AG within the research project “intermobil Region Dresden” funded by the German Federal Government, the ministry of Research and Education (BMBF). Project no.: 19 B 9907 A 8.

References

- [1] N.N., *Naissance d'un metro: The birth of a metro*, 1998. Special issue: La vie du rail & des transports.
- [2] Scholz, S., *Possibilities and limitations for a demand-dependent control of transport supply of commuter railways in medium-sized conurbations*. Ph.D. thesis, University of Technology, Dresden, 2005. Dissertation.
- [3] Scholz, S., Strobel, H. & Oettich, S., Demand-driven automated urban rapid rail transit: a new approach to the assessment of the operational efficiency. *APM '03*, 9th Int. Conf. on Automated People Movers, Singapore, 2003.
- [4] Goldberg, D.E., *Genetic Algorithms in Search, Optimisation and Machine Learning*. Addison-Wesley: Reading, Massachusetts, 1989.
- [5] Whitley, D., A genetic algorithm tutorial. *Statistics and Computing*, **4**, pp. 65–85, 1994.
- [6] Wegele, S. & Schnieder, E., Automated dispatching of train operations using genetic algorithms. *Computers in Railways IX*, WIT Press, pp. 775–784, 2004.
- [7] Kwan, R. & Mistry, P., A co-evolutionary algorithm for train timetabling. *2003 Congress on Evolutionary Computation*, IEEE, pp. 2142–2148, 2003.
- [8] Albrecht, T., Nutzbarmachung Genetischer Algorithmen für den automatisierten Entwurf flexibler S-Bahn-Betriebsregime. *at - Automatisierungstechnik*, **54(3)**, pp. 1–8, 2006.



An algorithm for train rescheduling using rescheduling pattern description language R

C. Hirai¹, N. Tomii¹, Y. Tashiro², S. Kondou¹ & A. Fujimori³

¹*Transport Information Technology Division,
Railway Technical Research Institute, Japan*

²*Kyushu Railway Company, Japan*

³*Hokkaido Railway Company, Japan*

Abstract

We propose an algorithm for automatic train rescheduling with a train rescheduling pattern language processing system. Intended for restoration from heavy train traffic disruption, our proposed algorithm has inherent abilities to make effective train rescheduling plans. While the previous algorithm tries to make a train rescheduling plan in small steps, the proposed one surveys the train timetable at first and applies “train rescheduling patterns” to prepare rescheduling plans. Applying actual train schedule data, we have confirmed that our algorithm works satisfactorily. For severe train traffic disruption caused by an accident, in particular, requiring the suspension of train operations for more than an hour, the algorithm is helpful for preparation of adequate rescheduling plans for practical application.

Keywords: train rescheduling, pattern description, train rescheduling pattern, train traffic disruption, train traffic control, framework of train rescheduling system.

1 Introduction

Train traffic is sometimes disrupted when accidents, natural disasters or technical problems occur on railway lines. In order to restore disrupted services, railways have continuously made a series of modifications to the current train schedules. Such a task is termed as train rescheduling [1, 2]. It is quite important for railway companies to prepare an adequate train rescheduling plan hereinafter referred to “the plan” whenever train traffic disrupted.



At present, expert train dispatchers are involved in train rescheduling. They constantly monitor daily train operations closely at their post. Once an accident that likely cause disruption of transport services occurs, train dispatchers collect information about it and prepare “the plan” based on their previous experiences and intuition. They take into account not only the future train traffic and the passenger density but also unforeseen problems.

It is required to introduce an algorithm for automatic train rescheduling. In order to maintain stable transport services, railway companies require a mechanism which ensures the quality of “the plan”. Recently computer systems adopted in operational services to assist train dispatchers in charge of train rescheduling, refer to [3, 4]. The systems, however, have a lack in a certain functions required to formulate “the plans” automatically.

In this paper, we propose an algorithm for automatic train rescheduling. Our algorithm based on a technique that applies experts’ knowledge as “train rescheduling patterns” hereinafter referred to “the patterns” to automatic computation of “the plan”. Some railway companies recognize specifiable parts of the experts’ knowledge as “the patterns” to restore disrupted train traffic effectively. In order to build a practical algorithm for automatic train rescheduling, a use of “the patterns” is recommendable.

On the other hand, we have to consider the fact that a mere use of “the patterns” is insufficient to prepare effective and practical plans. Due to uncertainty of occurrences of accidents, it is difficult to prepare complete train rescheduling patterns to suit to every situation. Such circumstance results in a need of a mechanism that modifies “the plan” prepared by only application of “the patterns”.

For the sake of constructing an automatic train rescheduling algorithm, we have to solve the following problems:

1. Establish a language to describe “the patterns”.
2. Develop a language processing system (an interpreter) to apply “the patterns”.
3. Construct a framework of an automatic train rescheduling system to modify the plan created by the interpreter.

We have solved above three problems, and implemented an algorithm for automatic train rescheduling. Applying actual train schedule data, we have confirmed that our algorithm works are satisfactory.

2 Train rescheduling pattern

2.1 Train rescheduling pattern

Some railway companies recognize specifiable parts of the experts’ knowledge as “the patterns” to restore effectively disrupted train traffic. Each train dispatcher has his own strategy in preparation of “the plan”, which causes variations in “the plans”. For example, under the same conditions, a dispatcher adopts a plan by which disrupted traffic restorable as immediately as possible while another dispatcher selects a plan that maintains sufficient traffic capacity.



In order to provide stable and punctual transportation services to public, every train dispatcher needs to own common and compatible strategies for train rescheduling. Railway companies accordingly have taken a step under which every train dispatcher utilizes “the patterns”.

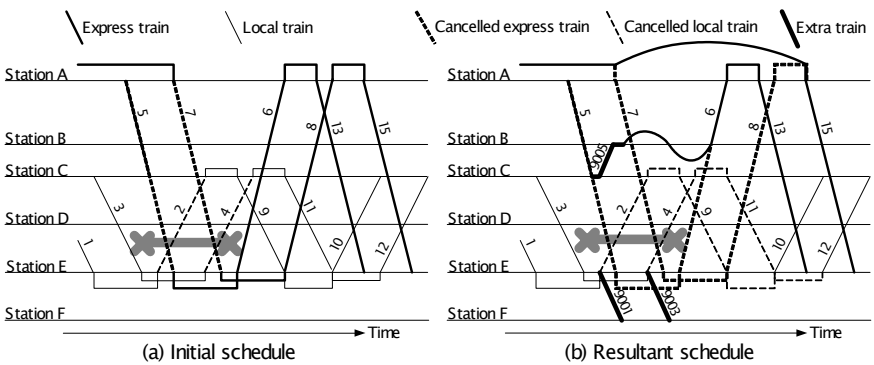


Figure 1: Example of train rescheduling.

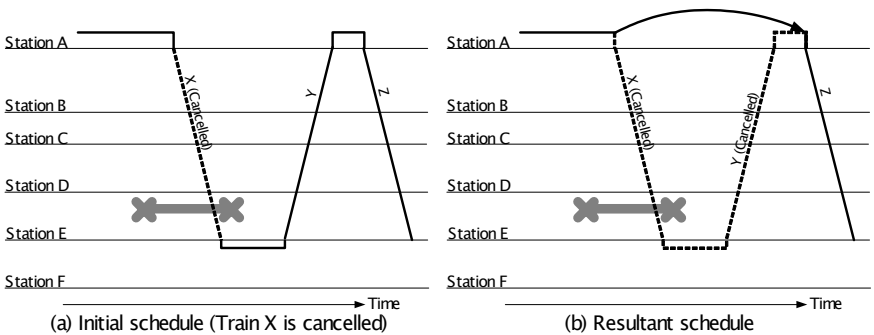


Figure 2: Train rescheduling pattern 1 (TRP-1).

2.2 Examples of applying train rescheduling patterns

Figure 1 indicates examples how “the patterns” are applicable when train traffic is disrupted. Figure 1(a) shows the initial schedule while Figure 1(b) shows the resultant schedule. A thick diagonal line indicates an express train, and a thin diagonal line shows a local train. Express trains operated between station-A and station-E; and local trains between station-C and station-E.

We assume that station-B and station-F have sufficient tracks to accommodate temporary train-sets. On the contrary, there is no extra track available at both station-C and station-E. Grey thick line indicates a location and an affected time interval of the accident where trains are inoperable. A broken diagonal line indicates a cancellation of the corresponding train.

Figure 1(a) illustrates initial schedule that includes train cancellations immediately after the accident. We assume that trains encountered with the

accident will be likely cancelled as initial operations. In Figure 1(a), the initially cancelled trains are train-5, train-7, train-2 and train-4.

Figure 1(b) shows a resultant schedule after applying the following patterns.

Figure 2 shows a train rescheduling pattern 1 (TRP-1). In case an express train-X is cancelled between station-A and station-E, the next express train-Y from station-E cancelled. A train-set for an express train-X waits at station-A, and operated as the next express train-Z.

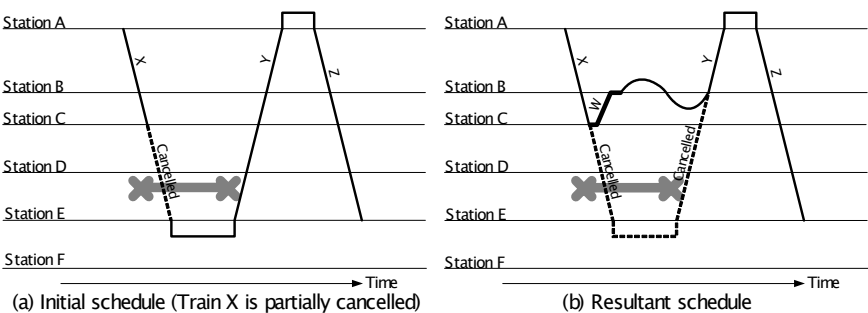


Figure 3: Train rescheduling pattern 2 (TRP-2).

Figure 3 shows a train rescheduling pattern 2 (TRP-2). In case an express train-X cancelled between station-C and station-E, the next express train-Y cancelled between station-E and station-B. An extra train-W is set from station-C to station-B. A train-set for train-X operated as a train-W and waits at station-B. After that, the train-set operated as train-Y from station-B to station-A.

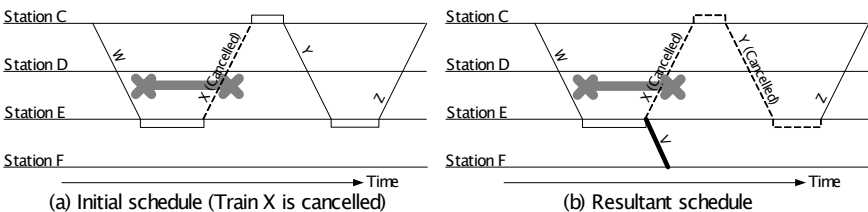


Figure 4: Train rescheduling pattern 3 (TRP-3).

Figure 4 shows a train rescheduling pattern 3 (TRP-3). In case a local train-X cancelled between station-E and station-C, the next local train-Y cancelled between station-C and station-E. An extra train-V is set from station-E to station-F. A train-set for train-X waits at station-F until the next operation is decided.

A resultant schedule in Figure 1(b) is derived from application of TRP-1, TRP-2 and TRP-3 to the corresponding cancellations respectively as shown in Figure 1(a).



2.3 Background of introducing train rescheduling patterns

Based on our hearing investigation of several railway companies, we can summarize reasons to make use of “the patterns” as follows:

- a. Provide stable transport services. It is possible to reduce variations due to different strategies taken by each train dispatcher in preparation train rescheduling tasks.
- b. Perform train rescheduling task smoothly. Train rescheduling operations involve many people; train dispatchers, drivers, conductors, station staff, etc. It is necessary for them to coordinate closely each other to avoid any failure. Familiarized with the detail of the plan in advance, every staff involved in the work is capable of fulfilling their activities with confidence.
- c. Avoid unforeseen problems. Skilled train dispatcher has expertise to avoid unpredictable trouble. Though it is not possible for them to explicitly explain the details of the trouble, in many occasions there are covert information not disclosed to the public. Hence we can incorporate the expertise into “the patterns” beforehand; train dispatchers are able to take advantage of expertise of other dispatchers.
- d. Realize train rescheduling operations that reflect a company policy of railways. It is possible to incorporate the policy into “the patterns” beforehand.

Considering the concept of railway companies mentioned as above, we conclude that a practical algorithm for automatic train rescheduling should utilize “the patterns”.

In addition, the utilization of “the patterns” creates a reduction of computing time. Our approach includes a simplified procedure; merely applying “the patterns” to the current schedule. In comparison with an algorithm that enables to prepare “the plans” successively from the beginning, our approach enables to construct an algorithm by which “the plans” are prepared within a short period.

2.4 Problems to be solved to introduce train rescheduling patterns

For the sake of constructing an automatic train rescheduling algorithm, we have to solve the following problems:

1. Establish a language to describe train rescheduling patterns: Up to now, “the patterns” are drawn in a natural language or drawn graphically as shown in Figure 2, 3 and 4. Therefore, it is necessary to establish a language that enables to represent them in a computer-recognizable form.
2. Develop a language processing system (an interpreter) to apply “the patterns”: A mechanism applying “the patterns” to the current train schedule is essential. In other words, the following functions are required.
 - A function to find a location where “the patterns” is applicable to the current schedule: in Figure 2, cancelled train-X should be found to apply TRP-1.



- A function to assign train names of the current train schedule to “the pattern”: in Figure 1(b) and Figure 4, it is necessary to assign train-2 and train-4 to train-X.
 - A function to output operation instructions: In Figure 1(b), the output should be cancellations of train-6, train-8, train-9 and train-11, settings of train-9005, train-9001 and train-9003, and related train connections.
3. Construct a framework of an automatic train rescheduling system to modify the plan created by the interpreter: The only use of “the patterns” is insufficient to prepare a practical rescheduling plan. It is difficult to prepare “the patterns” for every single case. Not all trains are applicable; therefore, a framework that modifies the results of applying “the patterns” is required.
- We have settled these problems. For each problem, our approach is as follows:
1. Introduce a pattern description language named “R”.
 2. Develop a pattern description language processing system named “R-interpreter”.
 3. Construct a framework for comprehensive train rescheduling system.

3 Pattern description language R and R-interpreter

We introduce a language named “R” to describe “the patterns”. In addition, we have implemented a language processing system named “R-interpreter” to interpret descriptive contents written in R.

We designate “the patterns” written in R “R-rule.” R-rule describes actions that executed when an event occurs. An event corresponds to a change of train schedule like a train cancellation.

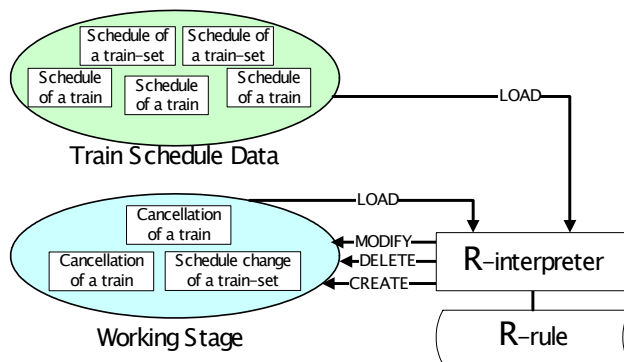


Figure 5: Operations of R-interpreter.

Figure 5 illustrates operations of R-interpreter. R-interpreter handles events at Working Stage based on R-rule, train schedule data and the current states of Working Stage. When R-interpreter is not able to find any of R-rule applicable to the current state, R-interpreter aborts its operations.



R-rule consists of three parts as below:

(1) (*for_each*) section

Describe an event that behaves a trigger of “the pattern”. R-interpreter searches the initial schedule for the trigger. In case that the trigger is found, R-interpreter try to check the conditions described in (*if_exists*) section.

(2) (*if_exists*) section

Describe conditions for actions.

(3) (*then*) section

Describe contents of actions.

```

begin Pattern_1
  for_each (
    運転中止 始端駅 = A, 終端駅 = E ] =: x in WS )
  if_exist (列車駅 列車番号 = x.列車番号, 駅 = A,
    列車種別 = 快速 運転方向 = 下り ] =: q,
    列車駅 列車番号 = x.列車番号, 駅 = E,
    列車種別 = 快速 運転方向 = 下り ] =: r,
    列車 列車番号 = 後運用列車(x.列車番号) ] =: w,
    列車駅 列車番号 = w.列車番号, 駅 = E,
    列車種別 = 快速 運転方向 = 上り ] =: s,
    列車駅 列車番号 = w.列車番号, 駅 = A,
    列車種別 = 快速 運転方向 = 上り ] =: t,
    列車 列車番号 = 前運用列車(x.列車番号) ] =: y,
    列車駅 列車番号 = y.列車番号, 駅 = A,
    列車種別 = 快速 運転方向 = 上り ] =: u,
    列車 列車番号 = 後運用列車(y.列車番号) ] =: z,
    列車駅 列車番号 = z.列車番号, 駅 = A,
    列車種別 = 快速 運転方向 = 下り ] =: v )
  then (generate 運転中止 in WS (
    列車番号 = y.列車番号, 始端駅 = E, 終端駅 = A )
    generate 車両運用変更 in WS (
    前運用列車番号 = y.列車番号,
    後運用列車番号 = z.列車番号 ) )
end Pattern_1

```

Figure 6: Description of TRP-1 in train rescheduling pattern language R.

Figure 6 shows an example that describes TRP-1 as shown in Figure 2. This description includes Japanese characters (Kanji and Kana characters), readily usable for Japanese train dispatchers. It is obviously possible to modify R-interpreter to deal with R-rule specified with only alphabetical characters.

The descriptive content in the (*for_each*) section in Figure 6 represents a cancellation of a train between station-A and station-E. This cancellation corresponds to that of train-X in Figure 2.

The description in the (*if_exist*) section represents conditions to be checked. In the (*if_exist*) section in Figure 6, the following conditions are specified;

- Existence of express train-Y that is operated from station-E to station-A (the same operating section as train-X's but the opposite direction).
- Existence of express train-Z that departs from station-A.



The (*then*) section in Figure 6 specifies actions of TRP-1. These actions correspond to the resultant schedule in Figure 2 are as follows:

- Cancellation of train-Y
- A train-set for train-X is operated as train-Z

4 Framework of Train rescheduling system

4.1 Framework of train rescheduling system

Train rescheduling is a complicated and large-scale task. A train rescheduling system, which supports train rescheduling tasks, should have a comprehensive structure.

Figure 7 indicates a framework for a comprehensive train rescheduling system. The framework consists of five subsystems: Train scheduling subsystem, Train-set rescheduling subsystem, Train rescheduling (in the narrow sense) subsystem, Crew rescheduling subsystem and Shunting rescheduling subsystem. Each of the subsystems exchanges relevant information and prepares each scheduling plan in cooperation with other subsystems.

It is necessary to distinguish operations that have different reasons. For example, a reason for a cancellation of a train is likely to be one of the following:

- In case where a train is express-service, a long delay loses its worth to arrive at a destination earlier.
- There is no train-set that can be assigned to the train.
- In order to restore the train traffic.

Therefore, Train scheduling subsystem, Train-set rescheduling subsystem, Train rescheduling (in the narrow sense) subsystem have each cancellation operation.

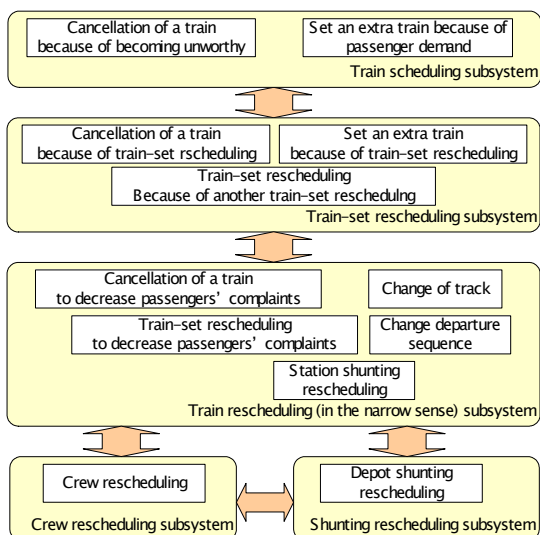


Figure 7: Framework of train-rescheduling system.



4.2 Train rescheduling algorithm

We propose an algorithm based on the framework shown in Figure 7. With the example of Figure 1, operations of each subsystem are indicated as follows:

4.2.1 Train scheduling subsystem

Train scheduling subsystem determines cancellations of trains and creation of extra trains. In Figure 1(b), the following cancellations are determined based on usage of R-interpreter as shown in Section 2.

- Between station-C and station-E, train-2, train-4, train-5, train-9 and train-11 are cancelled.
- Between station-B and station-E, train-6 is cancelled.
- Between station-A and station-E, train-7 and train-8 are cancelled.
- Train-9001, train-9003 and train-9005 are created.

4.2.2 Train-set rescheduling subsystem

Train-set rescheduling subsystem prepares a train-set operating rescheduling plan. In Figure 1(b), no train-set is assigned to train-10 and train-12. Then, train-sets for train-9001 and train-9003 assigned to train-10 and train-12 respectively. A train-set rescheduling algorithm [5] is applicable in this instance.

4.2.3 Train rescheduling (in the narrow sense) subsystem

Employed a train reschedule algorithm [6], this subsystem solves the detailed problems like track conflicts. The following rescheduling operations are dealt with:

Cancellation, Extra trains, Change of track, Change departure sequence, Change train-set operating schedule and etc.

4.2.4 Crew rescheduling subsystem

For the rescheduling plan, this subsystem enables to prepare a crew rescheduling plan with a crew rescheduling algorithm [7].

5 Results of experiments and evaluation of our algorithm

We have evaluated effectiveness of our train rescheduling algorithm using actual train schedule data. For experiments, we selected a line that has 20 trains for an hour. An accident that requires an hour train suspension assumed as indicated by a black thick line, therefore, trains heading for the downward direction are not able to go into the location for an hour.

Table 1: Evaluation values of algorithms.

Algorithm	Evaluation value
Without rescheduling (Figure 8(a))	2138
Existing algorithm [6]	686
Proposed algorithm (Figure 8(b))	331



As an index for evaluation, we employed “dissatisfaction index” proposed in [6]. This index indicates the level of passengers’ dissatisfaction, if the value is small; we regard the algorithm works satisfactory.

Table 1 indicates the result of the comparative evaluation. Because the value of the proposed algorithm is the smallest, we conclude that the proposed algorithm have made a better rescheduling plan than the existing algorithm. While the existing algorithm tries to prepare a train rescheduling plan tardily from the beginning, the proposed algorithm can utilize appropriate rescheduling policies given as R-rule in advance. Since R-rule contains expertise of skilful dispatcher, the proposed algorithm can prepare a practical rescheduling plan.

In addition, we can find obvious differences between disrupted schedule (no operation is executed) shown in Figure 8(a) and the resultant schedule shown in Figure 8(b).

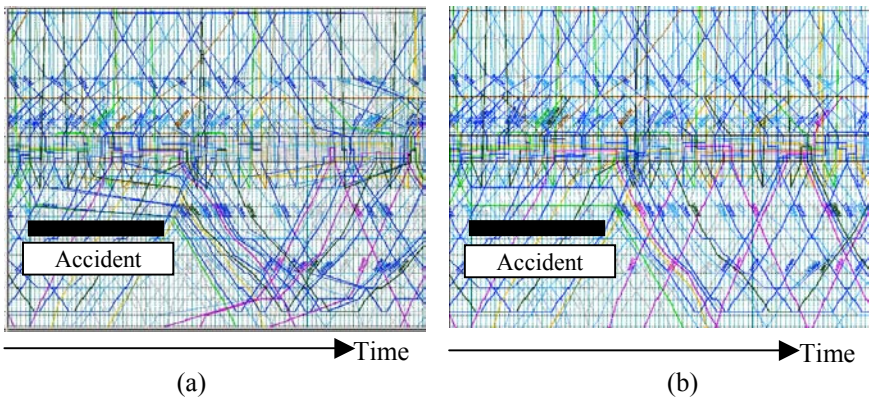


Figure 8: (a) Disrupted schedule and (b) Resultant schedule.

6 Conclusions

We proposed an algorithm for automatic train rescheduling with a train rescheduling pattern language processing system. Applying actual train schedule data, we have confirmed that our algorithm works satisfactory. Especially for heavy train traffic disruption, caused by an accident requiring a train suspension for more than an hour, the algorithm can produce a rescheduling plan suitable for practical use.

References

- [1] Tomii, N., *An Introduction to Railway Systems* (in Japanese), Kyoritsu Publishing, pp. 64-78, 2001.
- [2] Tomii, N., *Techniques to make train diagrams for punctuality* (in Japanese), Seizando Publishing, pp. 124-173, 2005.



- [3] Goodman, C. J. & Takagi, R., Dynamic re-scheduling of trains after disruption, *Computers in Railways IX*, eds. J. Allan, C. A. Brebbia, R. J. Hill, G. Sciutto & S. Sone, WIT Press, pp. 765-774, 2004.
- [4] Nakamura, T. & Ihara, K., The present situation and problems of train traffic control systems (in Japanese), *IEEJ Journal*, **124(5)**, pp.279-283, 2004.
- [5] Hirai, C. Tomii, N. & Tanabe, N., An algorithm for train set scheduling problems toward Cyber-Rail (in Japanese), *Transportation and Electric Railway*, TER-02-46, IEEJ, 2002
- [6] Tomii, N., Tashiro, Y., Tanabe, N., Hirai, C. & Muraki, K., Train rescheduling algorithm which minimizes passengers' dissatisfaction, *Proc. 18th Int. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE 2005)*, M. Ali & F. Esposito eds., LNCS 3533, Springer, pp. 829-838, 2005.
- [7] Fujimori, A. & Tomii, N., A Crew rescheduling algorithm based on combinatorial auction model (in Japanese), *Information Technology Letters*, **3**, LA-006, pp. 15-18, 2004.



This page intentionally left blank

Section 10

Timetable planning

This page intentionally left blank

State-of-the-art of railway operations research

I. A. Hansen

Delft University of Technology, The Netherlands

Abstract

The paper discusses the current state of research concerning the modelling of stochastic train operations in order to optimise the use of capacity by means of analytical methods and micro-simulation. Timetable quality is governed by precise running and realistic recovery times, as well as optimal headway and buffer times. Analytic (queueing) models and micro-simulation are used for estimation of waiting times, while combinatorial models and stability analysis by means of max-plus algebra technique are suitable for network timetable optimisation. Operations quality can be modeled by the estimation of (admitted) queue length and micro-simulation. A recently developed probabilistic model for the estimation of delay propagation in stations and junctions based on the stochastic variation of track occupancy times depending on conditional probability distributions of hinder by headway constraints and route conflicts is presented, while stochastic timetable optimisation may help to determine a more efficient distribution of running time margins and to increase punctuality in large networks.

Keywords: *timetabling, waiting time, delays, rescheduling, level of service.*

1 Introduction

So far, railway timetables are based on deterministic running, dwell and headway times between stations. These times are mostly scaled in minutes and refer to a virtual stopping point at the stations. Small variations of the service times are compensated by standard running time and dwell time supplements, as well as margins between the train paths. The amount of time supplements and margins applied for scheduling, however, is mainly based on rules of thumb, sometimes checked by simulation, and only seldom derived from statistical analysis of real-world operations data.



Stochastic analytical approaches, like queueing models [1–5] assume the train intervals and the service times to be independent random variables, which is questionable in case of periodic timetables with high frequencies. The variation of inter-arrival times and minimal headway times is modelled mostly on the basis of assumed random distributions for inter-arrival and service times. The estimated waiting time of a timetable is a function of the track occupancy and the coefficients of variation of the scheduled headway and service times. Waiting times generated by the stochastic variables of the timetable are clearly distinguished from estimated knock-on delays during operations.

Optimisation models based on linear programming [6], in general, apply deterministic variables for searching the optimal value of an objective function, like minimisation of overall running times in large complex networks, at given constraints e.g. minimal headway and transfer times between trains. The first known stochastic optimisation approach for estimating the robustness of railway timetables was recently presented by Vromans [7].

Micro-simulation models, like STRESI [8], RailSys [9], OpenTrack [10] or ATTPS [11] are used to estimate the effect of exogenous random primary delays on track occupation and consecutive delays of hindered trains. The induced primary delays are drawn from assumed or empirical distributions for a given timetable, track infrastructure and signal system. Kaminski [12] introduced a heuristic limit for the buffer time distribution at bottlenecks in order to compensate for 80% of the primary delays. Carey and Carville [13] presented a heuristic approach for solving conflicts between train paths and routes and simulation of random delays in order to test the reliability and robustness of timetable options.

This paper will discuss the influence of basic stochastic variables for timetable design and its impact on railway performance, as running and recover time, minimal time headway and punctuality. A new approach for the estimation of blocking times and buffer times at bottlenecks by means of blocking time ‘waves’ is presented. The paper concludes with the remaining issues for further research and development.

2 Timetable quality

The evaluation of timetable quality must reflect the requirements of the different stakeholders, as customers (passengers, shippers), infrastructure managers, train operating companies and public authorities (government, passenger transport executives). The timetable itself represents the transport supply, which should comply best with the market demand. There exist numerous criteria to evaluate the quality of railway services, which are more or less important for customers, infrastructure provider, train operators and public authorities. The quality of timetable design for a given traffic demand, rail infrastructure and train mix depend mainly on:

- Precise and realistic running times
- Sufficient but not too large recovery times



- Exact minimal headway times between different pairs of trains,
- Estimated waiting time.

2.1 Running and recovery times

Stochastic train running times were first modelled by Steckel [14]. He performed multiple simulations (# 400) with random input for different locomotive hauled trains and found Weibull distributed running times independent from the route and train characteristics with positive skew and parameter δ between 1.2 and 3.0. The peak factor (ratio of pure running time plus margin/ pure running time) varied strongly due to route conditions and train mass (Fig. 1). An approximate estimate of the superior value based on simple regression functions of the share of acceleration time is used to calculate the necessary running time margin of a certain route without intensive simulations.

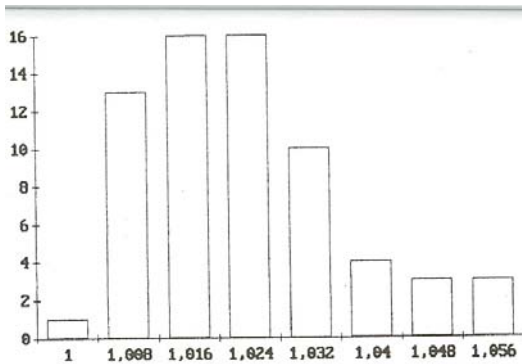


Figure 1: Frequency of running time peak factors of passenger trains (source: Steckel).

Empirical distributions of exact running times can be easily derived from track occupation and release data which are recorded and saved automatically by the existing signalling and safety system. As these data only contain the passing times per train number at main signals and these are distant from the stopping points at platforms the additional deceleration and acceleration time respectively from and to the last (first) main signal needs to be estimated taking into account the remaining distance between the last/first measuring point and the stop location, as well as the train characteristics (length, weight, power, standard deceleration/acceleration rate). At TU Delft the tools called TNV-Prepare and TNV-Filter [15] have been developed by which the recorded passing times at signals and insulation joints of any train can be filtered and the real running times and delays per train series be computed automatically according to their route, type and period of the day.

The statistical analysis of the precise running times per train line enables to determine the percentile of running times within a certain level of probability (Fig. 2). Furthermore, the true minimal running times per link can be revealed at



a precision of seconds by selecting only those trains that were delayed at the preceding departure station and experienced no hinder.

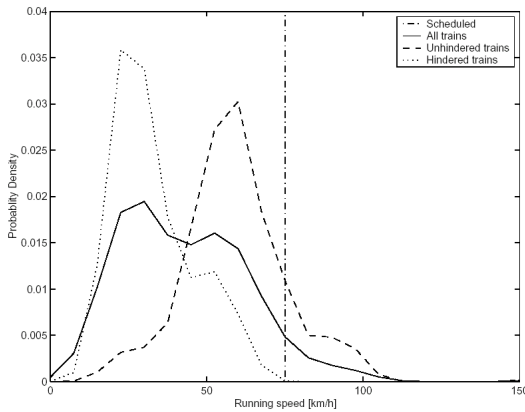


Figure 2: Distributions of running speed of a train series in case of (no) hinder North of the station The Hague HS (September 1999).

As the running times depend not only on the alignment, number of and dwell times at intermediate stops, type, length and weight of rolling stock operated, but also on the behaviour of the train drivers and the weather conditions, the specific performance distributions can be sorted and used as information for quality management and fed back to timetable designers and operations personnel.

The running time distributions per train line, thus, can be used for specifying the amount of recovery time to be applied in timetables instead of using standard supplements for all links and trains. The recovery times needed for achieving a certain punctuality level of the train arrivals based on departure delay distributions can be estimated numerically by means of convoluting the integrals of the departure delays and running times per link and train line [16].

2.2 Minimal headway and buffer times

In practice, timetable designers mostly apply standard mean minimal headway times (2 to 5 min) between train paths depending on the type of conflict and train sequence. The existing buffer times in a conventional graphical timetable which indicate only the train paths and headway times at a scale of minutes cannot be determined sufficiently, because these do not indicate the precise start and end of the capacity consumption according to the prevailing signalling and safety system.

The minimal time headway between two trains on a railway route is governed by blocking times of trains and its speed differences [17, 18]. The blocking time graph of a train represents the time instances that a train needs to run safely without hinder at design speed over a sequence of track sections. The scheduled train paths virtual as well as the track occupancy times are represented in a time-distance diagram and indicate clearly the remaining buffer times. Any overlap



between the blocking times of different trains clearly indicates a timetable conflict, which needs interaction in order to avoid a deceleration or stop of the following train.

The blocking times depend not only from the signal spacing and train length, but also from the actual train speed and deceleration rate. If the movement authority for the train at sight distance of the distant signal is given late because of insufficient headway to the preceding train, the following train would be decelerated automatically, which means an increase of the blocking time. From empirical analysis of train detection data at different Dutch railway stations follows that the mean speed of trains approaching to a station is about 10 to 20 % less than the design speed and may, in case of hinder, even be 50% less [19].

Furthermore, the scheduled dwell times are often exceeded during operations, which lead to an increase of the blocking time of the routes serving the platform tracks. The quality of timetable design would be enhanced considerably if the estimated blocking times reflect well the variation of train speed and dwell times in practical operations.

So far, the blocking times in timetables are assumed to be deterministic and estimated with a precision of seconds. The arrival and departure times of trains and the headway times between trains in most railway timetables, however, are determined currently with a precision of minutes due to rounding-up of the estimated running times and easy comprehension by the passengers. This practice includes hidden scheduled waiting times, which could be exploited if the determination of the arrival, departure and headway times (for in-company planning and operation purposes) was done at a higher precision in steps of 5 to 10 sec (according to current practice in Japanese railways).

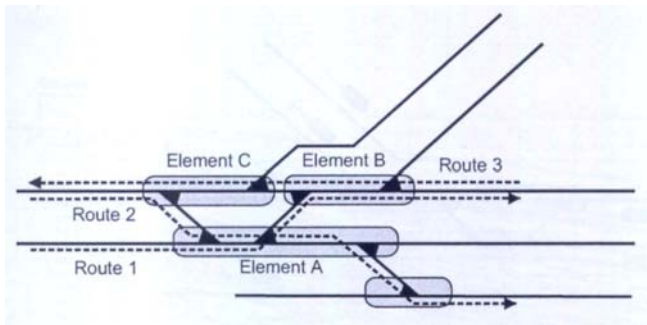


Figure 3: Division of an interlocking arrangement into route nodes [18].

For capacity estimation of heavily occupied routes in bottlenecks the track infrastructure needs to be subdivided in individual route nodes: the smallest track elements that can be used by one train at a time (Fig. 3). The variations of time headway and train speed at heavily occupied junctions may lead to knock-on delays and queueing of trains before network bottlenecks. On the routes approaching to level crossings close to main stations we observed a significant drop of the mean train speed by about 20 to 50% with regard to design speed

which means a clear reduction of infrastructure performance. This means the scheduled speed and headway times must correspond better to real operations and the train drivers need to be supported by dynamic speed advices in order to avoid hinder to and by other trains at junctions. In fact, the blocking times are stochastic, and not deterministic, and so are the remaining buffer times (Fig. 4).

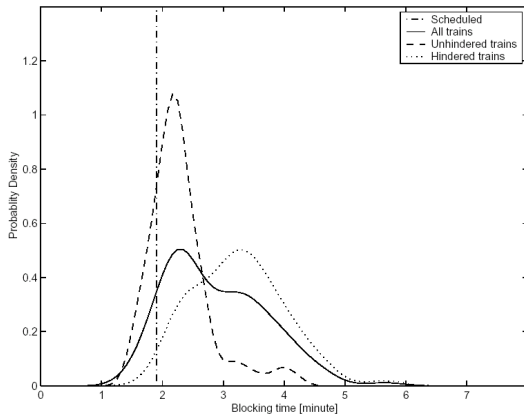


Figure 4: Distribution of blocking times of a train series at the station The Hague in case of (no) hinder (September 1999).

The probability density of the capacity consumption of the trains passing at bottlenecks could be represented by stochastic blocking time 'waves'. The range of blocking times covered by the stochastic waves of each train line indicates the extra use of capacity due to the variation of real operations in comparison with the deterministic blocking time estimated on the basis of the scheduled running time and possibly dwell time per train.

The total track occupancy and the remaining timetable slack of a given line is estimated finally by means of a virtual compression of the blocking time graphs according to the new UIC norm 406 capacity [20]. It recommends maximal track occupancy rates of 75% during peak hours and 60% a day for lines used in mixed traffic.

2.3 Waiting times

Any timetable comprises scheduled waiting times and generates, in practical operations, non-scheduled waiting times. Scheduled waiting times result from differences between the scheduled and desired travel, headway, departure or arrival times. The latter depend on the market demand, track possession times due to maintenance and conflicts between resulting different train graphs at bottlenecks, whereas non-scheduled waiting times can emerge from technical failures of track infrastructure or rolling stock, accidents, and train delays. Scheduled waiting times may occur randomly and its amount is used as indicator



of timetable quality, whereas non-scheduled waiting times are experienced as train delays during operations. The latter serve as indicator of operations quality.

The higher the reliability of tracks and rolling stock, the higher is their availability for operations and the practical capacity. The total waiting time during a certain time period, in general, increases exponentially with the number of trains operated. The challenge is to find an optimal track occupancy that enables a desired level of service, which can be defined by a certain amount of waiting time.

Schwanhäusser [1] developed already in 1974 a stochastic approach for the estimation of the mean queue length as function of the distribution of primary delays, buffer time, mean headway, train sequence and priority by a model of the type $M/D/1$. This led e.g. to an estimated mean buffer time of 1 min in case of a minimal headway of 2 min and all trains delayed (p. 59). As the assumed Poisson distribution of train arrivals near bottlenecks could not be confirmed later on, other queueing models of the type $G/G/1$ and $M/G/1$ have been developed by Wakob [2] and Hertel [3] respectively.

Hertel developed further an analytical approach for the determination of the optimal train intensity, which is situated between the minimal relative timetable sensitivity (partial derivative of mean waiting time to track occupancy) and the maximal traffic 'energy' (defined as product of train intensity and speed). The range of optimal use of track capacity as function of the waiting time, speed and 'energy' of the route, in general, lies at about 150 to 200 trains per day, while the mean waiting time per train would be up to about 10 min.

Wendler [4] extended in 1999 the existing two-train queueing model to a three-train model of type $G/G/1$ in order to estimate the impact of available time lags between the headway times on the main track of a junction for intermediate merging and crossing of trains. This approach is suitable for modelling the process of independent random requests for infrastructure capacity from different train operators and estimating the resulting waiting times due to timetable adaptation of train paths, headways and synchronization.

The relative errors of the estimated waiting times by the different queueing models with regard to the waiting times computed by micro-simulation in the case of the railway node Mannheim/Ludwigshafen as a function of track occupancy for exponential inter-arrival times are shown in Fig. 5. The 2-train queueing model of Wakob (thin line) generates exponentially increasing errors for individual 1 route nodes at higher occupancy rates of 0.7, whereas the Light-Traffic approximation of the 3-train model (bold line) of Wendler performs best up to 0.5 occupancy and the Heavy-Traffic approximation is the best analytical estimation for occupancy rates between 0.6 and 0.8.

Kaminski [12] analysed the blocking and buffer times in the existing timetable, as well as the recorded train delays for a large network of more than 2000 km in Germany and found a good fit of the buffer times with a negative-exponential distribution. The available buffer time was less than 1 min (standard minimal buffer time of Deutsche Bahn) in about 8% of the train sequences. He investigated the impact the length of buffer times that correspond to 80% of the recorded train delays at a number of stations by means of multiple simulation



while maintaining the train order. The results showed a mean consecutive delay per delayed train of about 30 sec and a reduction of 9% of the estimated hinder between trains. Thus, the expected operations quality could be improved by means of tailored buffer times per route section as function of the distribution of train delays.

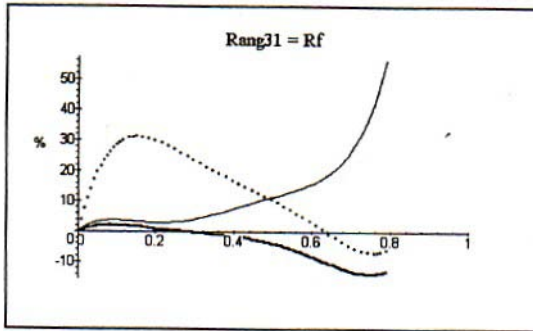


Figure 5: Relative error of waiting time estimation by different queueing models compared to micro-simulation (source: Wendler).

Microscopic simulation tools like STRESI [8], RailSys [9] OpenTrack [10] respectively estimate the total waiting times of lines and networks by means of computing the impact of multiple drawing of random primary delays from a predefined distribution on total waiting times. The virtual hinder generated by modification of the scheduled blocking time graphs is transformed automatically into suitable conflict resolution measures (bending of train paths, increasing of dwell times, change of train order and routing), while the impact is estimated by the distribution of the total amount of delays, the kind and percentage of hindered trains and the punctuality level.

The timetable slack of interconnected lines in networks with periodic timetable depends not only on the buffer times of the individual lines, but also on the buffer times between the scheduled train arrivals and departures at the transfer stations. The following constraints for the design of an integrated network timetable hold:

- The round trip time of a line must be integer multiple of the headway,
- The travel time between the nodes (sum of the running and dwell times) must be an integer multiple of half of the cycle time (mostly 60 min),
- The scheduled arrival and departure times of interconnected lines at the nodes must overlap sufficiently,
- The sum of the travel times on a circuit between three nodes must be an integer multiple of the cycle time and of the time headway.

As these constraints lead to a great complexity of periodic timetables in large networks with many nodes and lines, combinatorial methods by means of mathematical programming have been developed to find an optimal solution for integrated network timetables [6]. The network timetable problem is known as NP-complete and requires the development of intelligent algorithms to solve

within a reasonable computation time – if a solution for the given constraints exists – and/or to relax some timetable constraints in order to achieve an approximate solution. The predefined minimal headway and transfer time constraints applied in combinatorial models don't guarantee its practical feasibility in complex junctions and need to be proven by a detailed analysis of blocking times.

The existing buffer times and its distribution in periodic timetables of large networks can be calculated analytically by means of the (max, +) algebra approach, which is characterised by a transformation of the timetable constraints into a set of simple recursive linear equations [21]. This approach enables to identify the critical circuits within the network and to calculate the minimal cycle time of the timetable, as well as the existing timetable slack. A tool for the automatic transformation of conventional periodic network timetable constraints into a (max,+) state matrix of the travel times, minimal headway times and transfer times, for critical circuit and recovery time analysis has been developed. The impact of an increase or decrease of travel times and buffer times on the timetable slack and on the location of the critical circuit of periodic network timetable can be estimated rapidly, as well as the propagation of train delays in the network.

3 Operations quality

The operations quality can be assessed by means of different analytical approaches and micro-simulation. A sufficient level of performance requires, of course, that the basic timetable is feasible and the scheduled running times, dwell times and headways are realistic, which means the timetable must not contain any conflict between train paths proven by the absence of overlaps between the blocking time graphs. The impact of primary delays on the robustness of train services is expressed e.g. by the queue length, amount of knock-on delays, the number of involved trains, links and stations, as well as the fading-out time. Given a certain distribution of primary delays, the consecutive delays may not grow steadily due to insufficient buffer times. The scheduled times and track occupancy, thus, must allow for recovery.

3.1 Queueing model of RWTH Aachen

Schwanhäusser [1] defined the smoothness of operation as the share of unscheduled waiting trains at any location in order to proceed. He developed an analytical approach for estimating the consecutive delays in case of hinder based on queueing theory. The admitted level of hinder during operations was defined empirically by a maximal queue length (equal the maximum total consecutive delays per day), which is given by a simple exponential equation as function of the share of passenger trains:

$$L_W = \alpha \cdot e^{-\beta \cdot pP}$$

L_W : queue length [-]

α, β : coefficients



pP : probability of passenger trains

$$L_w = 0.257 \cdot e^{-1.3 pP} \quad (\text{calibrated function for Deutsche Bahn AG}).$$

Three different levels of quality of operations (0.5=very good, 1.0=satisfactory, 1.5=unsatisfactory) are defined by the ratio of the estimated length of queue divided by the maximum admitted length of queue. The standard maximum amount of knock-on delays to be expected due to operation of a certain number of trains admitted by Deutsche Bahn is 130 min up to 300 min per day depending on the share of passenger trains. These threshold values of maximum waiting time, however, look somewhat arbitrary and seem to need further empirical validation. The estimation of the scheduled and unscheduled waiting times as a measure of operations quality of a given timetable, infrastructure and signalling system is implemented in the software tool ANKE developed at RWTH Aachen [22].

3.2 New probabilistic model of TU Delft

The weakness of queueing models in case of periodic timetables can be overcome by an advanced probabilistic model of the process times of each train line (running or arrival time, dwell time, departure time) as stochastic variables, while the hinder due to signalling and route conflicts by delayed trains and its priority are taken into account [16].

The approach is based on the stochastic variations of track occupancy times e.g. the variations of train speed in case of different signal aspects with conditional probability distributions. Distinction is made between

- hindrance of approaching and departing train respectively by late release of preceding block or route section and
- Hindrance of approaching and departing train respectively by late release of crossing (merging) of routes.

The model is solved on the basis of a numerical approximation of the Stieltjes convolution of individual independent distributions.

The delay propagation to a train approaching to a station is modelled depending on the actual signal aspects, which may force the train to decelerate or even to stop. The arrival time distribution can be estimated by the probability that the train experiences more or less hinder and the conditional probability of the arrival time in case of each condition. This distribution is formulated as,

$$F_{A_i^p}(x) = P[A_i^p \leq x] = \sum_{h=1}^3 P[(A_i^p \leq x) | (\text{Condition } h)] \cdot P[\text{Condition } h]$$

The departure of a train at station is modelled analogously as function of the probability of dwell time extension and route conflicts and the distribution of running times until release of the routes. For more details of the analytical model see [16]. The calculation of the convolution is done numerically by applying the composite trapezoidal rule. The proposed model has been validated by means of train detection data recorded at The Hague HS station. To validate the model, we obtained the required input distributions on the basis of empirical track occupation and release data.



3.3 Stochastic timetable optimisation model of EU Rotterdam

Vromans [7] developed recently a timetabling model based on a Periodic Event Scheduling Problem (PESP) approach and combined with a reliability evaluation by means of simulation with a fixed order of events. The delay propagation is calculated by performing a number of consecutive runs of the scheduled trains, while these are perturbed by exogenous random disturbances and measuring the resulting arrival delays. The objective of the linear programming model is to minimise the mean arrival delay over all runs by means of an optimal allocation of the running time supplements along the line. If the standard 7% running time supplement of the Dutch Railways was distributed optimally instead of linearly, the punctuality of the trains operated in the investigated cross-country corridor Haarlem-Maastricht would be increased significantly.

3.4 Micro-simulation

The propagation of non-scheduled waiting times (delays) in large networks of interconnected lines can be modelled also by micro-simulation. Watson [23] compared the characteristics of current commercial simulation tools and concluded that ‘signal berth’ level tools are necessary for stochastic simulation. Simulation tools mostly require interaction of the user in case of conflicts between blocking time graphs or the application of a predefined automatic conflict resolution strategy. For the evaluation of different dispatching measurements on the stability margin and the location of network bottlenecks in case of disturbance different options can be computed. The analysis of individual link and train dependent recovery times would allow a variety of experiments to estimate the robustness of different re-scheduling options (re-timing, re-ordering, re-routing). The effectiveness of different conflict resolution strategies can be evaluated on the basis of the amount of disturbance (primary delay, knock-on delays, punctuality) and fading-out time.

4 Conclusions

The current methods and tools for railway timetable design enable a high precision of the estimated travel times, headway times and time margins in order to achieve high-quality conflict-free deterministic timetables for lines and networks. Queueing models and micro-simulation tools have been applied successfully to improve timetable quality and to estimate stochastic train delay propagation on lines, in stations and networks. The key for high-quality timetabling is a precise estimation of blocking times based on realistic running, dwell and headway times taking into account the signal spacing and train processing at critical route nodes and platform tracks. Statistical analysis of empirical track occupation and release data in Dutch stations has revealed that the trains mostly leave later and operate at speeds less than scheduled due to drivers’ and conductors’ behaviour, lack of movement authority and routes still locked for other delayed trains. Deterministic models for estimation of the stability or optimisation of network timetables, like max-plus algebra technique



or linear programming, may not represent real operations and lead to suboptimal solutions.

As queueing and simulation models for estimation of unscheduled delays in daily traffic, too, do not reflect sufficiently the impact of speed variations and behaviour of railway staff and may be based on invalid input distributions, new stochastic analytical approaches have been developed recently in the Netherlands. The probabilistic model of TU Delft enables an estimation of the survival rate of knock-on delays at platform tracks and junctions based on the distributions of the involved train pairs. It allows an estimation of the admitted maximal number of trains and distribution of consecutive delays at a given primary delay and level of punctuality. The stochastic timetable optimisation model of EU Rotterdam helps to find an optimal distribution of running time supplements along a line that may increase punctuality significantly.

Acknowledgements

This publication is a result of the research project 'Modelling and analysis of operations in railway networks: the influence of stochasticity' supported by the Dutch research foundation STW and Transport Research Centre of Delft University of Technology.

References

- [1] Schwanhäusser, W. (1974) "Die Besetzung der Pufferzeiten im Fahrplangefüge der Eisenbahn", *Veröffentlichungen verkehrswissenschaftl. Institut RWTH Aachen*, PhD thesis, H. 20
- [2] Wakob, H. (1985) "Ableitung eines generellen Wartemodells zur Ermittlung der planmäßigen Wartezeiten im Eisenbahnbetrieb unter besonderer Berücksichtigung der Aspekte Leistungsfähigkeit und Anlagenbelastung", *Veröffentl. verkehrswissenschaftl. Institut RWTH Aachen*, PhD thesis, H.
- [3] Hertel, G. (1992) "Die maximale Verkehrsleistung und die minimale Fahrplanempfindlichkeit auf Eisenbahnstrecken", *Eisenbahntechn. Rundschau ETR*, 41, H. 10, 665-672
- [4] Wendler, E. (1999) "Analytische Berechnung der planmässigen Wartezeiten bei asynchroner Fahrplankonstruktion", *Veröffentl. verkehrswissenschaftl. Institut RWTH Aachen*, PhD thesis, H. 55
- [5] Huisman, T., Boucherie, R.J., van Dijk, N.M. (2002) "A solvable queueing network model for railway networks and its validation and applications for the Netherlands", *EJOR*, 142, 30-51
- [6] Cordeau, J.F., Toth, P., Vigo, D. (1998) "A survey of Optimization Models for Train Routing and Scheduling", *Transportation Science*, Vol 32, (4), 380-404
- [7] Vromans, M.J.C.M. (2005) "Reliability of Railway Systems", PhD thesis, Erasmus University Rotterdam



- [8] Schultze, K. (1985) "Modell für die asynchrone Simulation des Betriebes in Teilen des Eisenbahnnetzes", *Veröffentl. verkehrswissenschaftl. Institut RWTH Aachen*, PhD thesis, H. 38
- [9] Radtke, A., Hauptmann, D. (2004) "Automated planning of timetables in large railway networks using a microscopic basis and railway simulation techniques", *Computers in Railways IX*, 615-625
- [10] Nash, A., Huerlimann, D. (2004) "Railroad simulation using OpenTrack", *Computers in Railways IX*, 45-54
- [11] Carey, M., Carville S. (2000) "Testing schedule performance and reliability for train stations", *Journal of Operational Research Society*, Vol. 51, 666-682
- [12] Kaminsky, R. (2001) "Pufferzeiten in Netzen des spurgeführten Verkehrs in Abhängigkeit von Zugfolge und Infrastruktur", *Wissenschaftl. Arbeiten Institut für Verkehrswesen, Eisenbahnbau und -betrieb Universität Hannover*, H. 56, PhD thesis, Hestra: Darmstadt
- [13] Carey, M., Carville S. (2003) "Scheduling and platforming trains at busy complex stations", *Transportation Research, Part A*, Vol. 37, 195-224
- [14] Steckel, J. (1991) "Strategische Optionen für die zufällige Fahrzeit im Eisenbahnbetrieb" (Strategic options for the random running time in railway operations), PhD thesis, Hochschule für Verkehrswesen "Friedrich List", Dresden
- [15] Goverde, R.M.P., Hansen, I.A. (2002) "TNV-Prepare: Analysis of Dutch Railway Operations Based on Train Detection Data", *Computers in Railways VII*, 779-788
- [16] Yuan, J., Hansen, I.A. (2005) "Optimizing Capacity Utilization of Stations by Forecasting Knock-on Train Delays", 1st Int. Seminar on Railway Operations Modelling and Analysis, Delft
- [17] Wendler, E. (2001) "Quality management in the operation planning process by means of harmonized modeling", *WCRR 25-29 Nov. Cologne*, paper no. 302
- [18] Pahl, J. (2002) "*Railway Operation and Control*", VTD Publishing
- [19] Yuan, J., Goverde, R.M.P., Hansen, I.A. (2002) "Propagation of train delays in stations", in: Allan, J. et al. (ed.) *Computers in Railways VIII*, WIT: Southampton, Boston, 975-984
- [20] UIC Union International des Chemins de Fer (2003) "Leaflet 405-1 Capacity"
- [21] Goverde, R.M.P. (2005) "Punctuality of Railway Operations and Timetable Stability Analysis", PhD thesis, TRAIL Thesis Series T2005/10, Delft
- [22] Vakhtel, S. (2002) "Rechnerunterstützte analytische Ermittlung der Kapazität von Eisenbahnnetzen", *Veröffentl. verkehrswissenschaftl. Institut RWTH Aachen*, PhD thesis, H. 59
- [23] Watson, R. (2005) "Using stochastic simulation to predict timetable performance – status and developments in the UK", 1st Int. Seminar on Railway Operations Modelling and Analysis, Delft



This page intentionally left blank

Timetable management and operational simulation: methodology and perspectives

A. Radtke

*Institut für Verkehrswesen, Eisenbahnbau und –betrieb,
University of Hannover, Germany*

Abstract

This article summarizes the fundamental basics for the timetable management of the future. The advantages of a microscopic database and some options for its smart utilisation are shown. After describing some important areas of application of the microscopic methodology, some practical experiences are discussed. The “Seven-Step-Model” is a first attempt to provide a methodology for using microscopic models. The description of the process highlights the iterative approach. The presented methodology can be used for the evaluation of the effort and time requirements for various planning tasks. The article closes with an outlook and a discussion about solving the problems of the planning task in the future.

Keywords: methodology of planning of railway operation, timetable construction, operational railway simulation.

1 Introduction

The timetable management is a key process for any railway company (RC) or train operating company (TOC). The general goal is a better timetable construction and the provision of a reliable train operation. This goal can be achieved with the RailSys simulation software. Over the last few years, the technology has been established that allows for the handling of very large networks and more than 10,000 trains in one single RailSys model [1, 2]. Therefore, it is now possible to determine the effects of conflicts between trains, or the propagation of delays in networks of any size and complexity.

The operational simulation of railways supports this process. Predicting the robustness of a timetable is a critical issue for the future. With the help of



RailSys both RC and TOC will be in the position to check a timetable before the timetable is operated in reality. A “simulation in reality” of a new timetable may contain a high risk for RC or TOC in the future due to penalties for poor On-Time-Running performance (OTR), or by jeopardising patronage or business from freight. In conclusion, the operational simulation of new timetables may become a “must” in the future and should be integrated into the workflow of the timetable production.

This article describes the fundamental aspects, the areas of application and some practical experiences. An outlook about some changes of processes and the latest developments in this context are given.

2 Fundamentals

The fundamental aspects of the state of the art timetable management and operational simulation are exemplified in the following paragraphs:

2.1 Microscopic infrastructure model

A node-link model describes the real infrastructure in a microscopic infrastructure model. All relevant information for planning a reliable timetable can be contained in such a model. The microscopic infrastructure model combines track information, such as speed, gradient, or radius, with the signalling system (type, blocks) and some operational information like routes or timing points.

Microscopic models are used for the operational planning of networks, lines and junctions. The modelling of infrastructure and operation in networks helps to avoid suboptimal solutions. The interrelation of the infrastructure, the rolling stock and the timetable should only be assessed on a network level. The complexity includes various signalling systems, different vehicle dynamics and train scheduling rules as well as additional operational conditions. For example, the identification and evaluation of reliable capacity indicators for railway lines require the inclusion of at least the adjacent junctions. It became apparent during the practical application of simulation methods that all branch lines of those junctions should be considered as well.

A simplified example in Figure 1 illustrates a transition between different infrastructure models. In a “Bottom Up” approach, it is evident that all macroscopic models can be derived in a straightforward manner from microscopic models. The “Top Down” approach can be used for generating microscopic models whose accuracy of detail depends on predefined assumptions. The “mesoscopic infrastructure model” has been placed in between both models. The models can be generated for specific tasks such as the simplified railway simulation of complex networks to answer some strategic questions. “Multi scale” models contain both areas that are modelled on a microscopic level and areas that are modelled on a macroscopic level. The advantage of such models is the minimisation of effort for modelling aspects of a complex problem that are of insignificant relevance for the over-all outcome of



the investigation (e.g. the individual operation of shunting yards or vehicle depots for the planning of a network wide timetable).

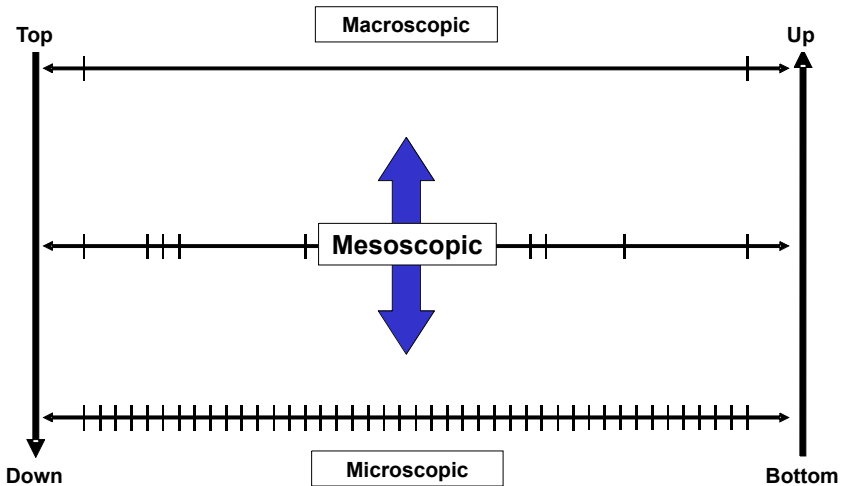


Figure 1: Multi scale infrastructure models.

It might be necessary to increase the level of detail of multi scaled infrastructure models for specific applications. This can be achieved without any problems, if the correct microscopic data is available. The further discussion in this article refers only to the microscopic models.

2.2 Exact running time calculation

The exact calculation of running times is essential for the correct planning of timetables. It should be possible to immediately recalculate running times at any time in microscopic models. For example, rerouting a train into another platform or changing the weight of a train may lead to significantly different running times. The variation of the running time may result in conflicts between trains, which need to be considered during the timetabling process. Using pre-processed sectional running times is, from our point of view, only a valid option if a microscopic infrastructure network for the timetable planning process is not available.

2.3 Block occupation and conflict detection

The microscopic description of the infrastructure model allows for the identification of conflicts and calculation of headways in the network, without regard to if only lines, junctions, or whole networks are considered. Figure 2 illustrates the principle: the horizontal axis is used for the geographical reference (signals, stations); the vertical axis represents the time. The “gradient” of the individual time-distance graphs reflects the speed of the trains. The block occupation boxes highlight the track occupation by a train. The resulting



occupation depends on the signalling and block system. Additionally the minimum headways can be derived from the illustration, which provide a first indication of the track capacity.

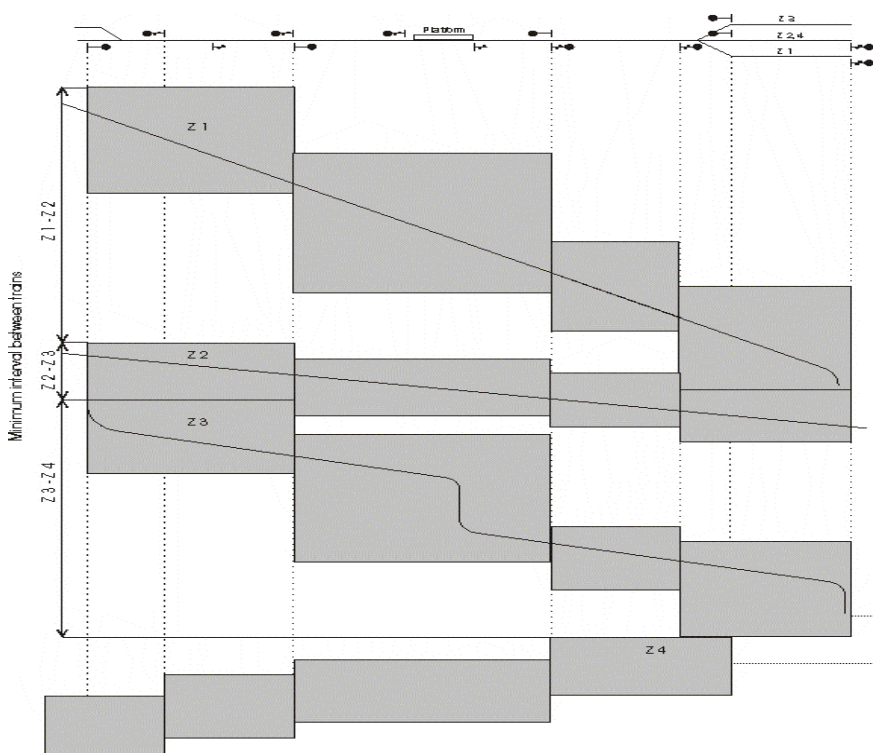


Figure 2: Block occupation [3].

2.4 Operational synchronous simulation

A synchronous simulation can be described as a dynamic process, in which all trains are running according to their schedule. Therefore, it is possible that some trains influence other trains, which results in delays or dispatching actions. In contrast, asynchronous methods handle all trains according to a hierarchy and assign trains one after the other into the model starting with high priority trains and ending with low priority trains. It is evident and has been proven that only the synchronous simulation models can provide indicators such as OTR performance reliably or accurately describes the delay propagation in railway networks. OTR values are very important for the design and operation of business rules and penalty regimes between RC and TOC.

The synchronous simulation can be used for two types of investigation. The first type is the simulation of a timetable to investigate the effects of inherent conflicts spreading throughout a network. During the timetable simulation there will be no introduction of additional delays, for example dwell time extensions.



The second part is the operational simulation (formerly called stochastic or multiple simulation). The operational simulation includes the introduction of additional delays according to statistical rules. In general, between 50 and 200 timetable variations with different delays must be simulated to obtain a sufficient statistical basis for the evaluation. Both the timetable and operational simulation use powerful dispatching algorithms to reflect the real time dispatching [4].

The following list summarises the most important elements and advantages of the synchronous simulation, which result in a reliable representation of real world conditions and operation:

- All trains are simulated in parallel to realistically reflect the interaction between trains. Included in the simulation are deceleration and acceleration of trains due to conflicts between one another,
- Consideration of external influences such as passenger amount, driver response to signalling, or technical defects of signals, points, or rolling stock,
- Utilisation of dispatching algorithms for solving operational problems,
- Recording of all simulation data for subsequent evaluation.

Determination of base characteristics for the evaluation of the performance		
Determination of base characteristics		measured as/in
1. Feasibility of a path	1.1 in one operating day	Number of paths [Quantity]
	1.2 at a desired time	Difference [hh:mm:ss]
2. Transport	2.1 time	Transportation time [hh:mm:ss]
	2.2 route	Transportation route [km]
	2.3 (speed)	Transportation speed [km/h]
3. Delay development	Average delay	[hh: mm:ss]
4. On-time	Delay frequency in classes	[mm:ss]
5. Connections	Connections reached	[Number of days], [%]
6. Vehicle rostering	Train reversals (circulation)	[Number of days], [%]



Figure 3: Selection of key indicators for operational simulation.

2.5 Evaluation of and key indicators for operational synchronous simulation

The evaluation of the results of the operational simulation provides essential information about the advantages or disadvantages of a timetable or



infrastructure concept. Some of the most important performance characteristics are displayed in Figure 3. The framework in which these key indicators fit may be different from RC to RC. However, the main point is to prove the reliability of the railway operation.

It will be possible in the future to manage timetables in a more complex environment with the help of those key indicators. Operational requirements such as vehicle rostering or the possibility to keep connections for passengers will have an increased influence over the timetable construction in the future.

3 Areas of application

In general, we can distinguish between the following areas of application for the microscopic timetable simulation. Of course, there are tasks that cover more than one area.

3.1 Long Term Planning (LTP)

The LTP deals with the establishment of timetable concepts that are based on the demand prognosis for freight and passenger traffic for a time horizon of five and more years. The amalgamation of macroscopic and microscopic data models has recently been achieved by linking the Network Modelling macroscopic model (NEMO) with the RailSys microscopic model in Austria for the LTP. The operational concepts that are developed by NEMO are transformed into a timetable on a microscopic railway network. The transformation is performed with the RailSys slot search engine that is described in [4].

3.2 Timetable design

The core business of an RC is to provide the rail infrastructure and the planning competence to schedule trains. The scheduling of trains has to pay regards to the system immanent railborne character of railways traffic, which significantly restricts the flexibility of movements. Therefore, the timetable design is one of the most important tasks to be considered. The sequence of trains and the resulting timely order is predetermined by the timetable design and can hardly be altered at a later stage.

In general, all RC have to solve the problem of providing a maximum number of train paths (we may call it in a simplified way here “capacity”) on a given infrastructure for a minimum of operational cost. In addition, political demands and requirements of TOC have to be taken into account, for example to operate passenger trains in rural areas even without gaining profit or covering the operational costs.

The timetable design itself is a highly dynamic process. It takes in general up to 24 months to design a timetable. The process involves various steps from the first draft to a working timetable. The requirements include the feeding of internal and public information systems or the provision of print outs for the operation. There are various iterations necessary to work out an “optimal” solution between all parties involved (RC, TOC and the market). In addition, the



available budgets (often public money) will influence this process; competitors like the air sector have to be considered when planning a timetable.

3.3 Short Term Planning (STP) and Ad-Hoc-Dispatching

The simulation of railway operation can provide valuable assistance and support for both STP and ad-hoc dispatching.

STP may be required for operating a scheduled timetable with a shortage of suitable vehicles, or for planning urgent possessions for track work. It could be necessary to solve conflicts during the real time operation due to variations of train length, weight or locomotive supply. Temporary speed restrictions must also be catered for. Other specific situations may require urgent dispatching actions, for example bad weather or infrastructure failures. In all these instances a methodology that is based on the above-described fundamentals is suitable for maintaining a reliable operation. Software such as RailSys, which clearly reflects those principles, is a most valuable tool.

3.4 Vehicle allocation

Like most of the afore mentioned tasks, the allocation of rolling stock is an essential part of the planning process. However, today there are in general two individual departments in railway organisations who deal separately with this issue. In general, it is a sequential process. The timetable design is followed by the vehicle allocation. RailSys offers an interactive interface to the Dispo software. The interface facilitates the export of the infrastructure network, the timetable, and other operational information that is essential for performing the vehicle allocation.

3.5 Support for franchise bidding

The preparation of franchise biddings in the UK and other countries with liberated railway networks is becoming an increasingly frequent task. Therefore, the process of timetable design and vehicle allocation needs to be performed quasi simultaneously. In the UK, it is possible to obtain a microscopic RailSys network model and existing timetable information from Network Rail under certain conditions. This fundamental information enables each franchise consortium to plan an exact timetable with correct running time calculation and conflict detection. Furthermore it is possible to predict the OTR of the planned timetable and therefore to reduce the risk of paying penalties for not achieving the required OTR performance.

3.6 Possession planning

The planning of possessions for the maintenance of existing railway networks will become a more important task. This is particularly true due to the decreasing budgets for new railway infrastructure in European countries with already



established railway networks. It will be essential for the RC to provide reliable information to the TOC in time about alterations of the timetable due to changes in running times, which might be a consequence of track blockings or temporary speed restrictions.

For all these tasks, the synchronous simulation of microscopic infrastructure models can be seen as a powerful instrument for the control of the proposed plans. Furthermore, the slot search engine can be used during all planning stages. It provides additional support to the planner when creating new timetables or altering existing train runs.

4 Practical experiences

The Seven-Step-Model is derived from the experience that has been gained through the practical application of microscopic simulation models. This approach is the first step towards a comprehensive methodology on working with microscopic models. The principles are not limited to RailSys only; all synchronous simulation tools may be embedded in that methodology. In the following list the main steps are defined:

Step	Description
I	Definition of the project type
II	Data acquisition and essential elements of the model
III	Validation of the base model
IV	Timetable design
V	Timetable studies and slot management
VI	Operational simulation
VII	Evaluation and presentation of results

Figure 4 shows that it should be considered at the beginning of a project if software enhancements are useful or necessary. If software enhancements are required, special care should be taken for the specification, programming and testing of any software enhancements. However, software amendments are very rarely required.

The model calibration and validation takes place during steps III and VI. Those activities make the highest demands on user input and know how. At these points, the modelling techniques may have a certain influence on the workflow or even on the results. Modellers with experience cannot be replaced by software!

On the other hand, the modelling know how can be learned in a reasonable amount of time when the Seven-Step-Model is applied and projects are structured very well. A structured approach and knowledge transfer guarantee reliable results. The definition of rules for the modelling process or the application of a “best practise” can significantly support the transfer of knowledge.



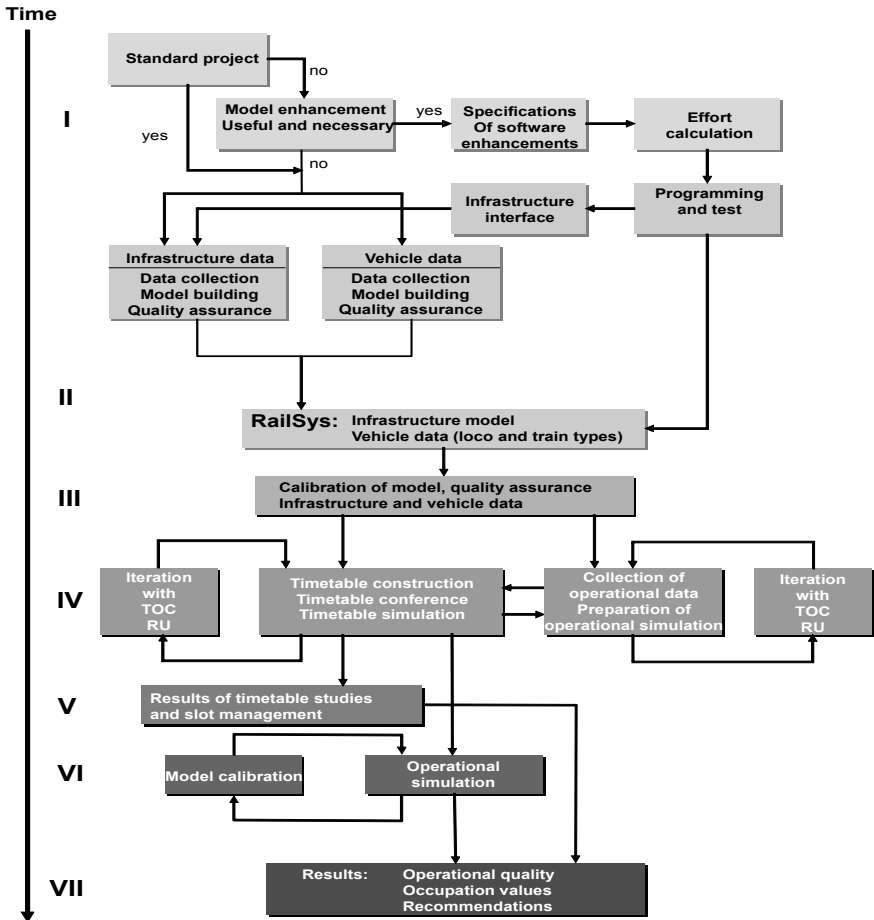


Figure 4: Seven-Step-Model [5].

5 Perspectives of the planning process

The future in the planning of railways may be illustrated in the following example. The illustration should be considered only as schematic. It does not illustrate or handle any details. This example considers only four activities or peaks during the planning process:

- Long Term Planning
- Timetable Construction
- Vehicle Allocation
- Possession planning

In general, for each of the four activities we can assume as a worst case that different teams are working with different tools on different databases.



Furthermore, those teams may be assigned to different business units within one company or even in different companies. This situation may of course cause many redundancies in maintaining databases and organizing the workflow. The amount of resources for administration might be relatively high in order to manage this situation.

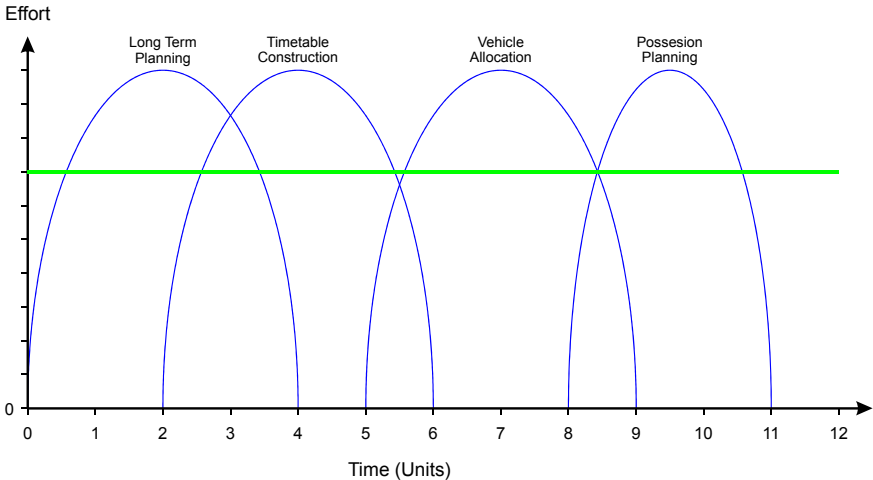


Figure 5: Future planning process.

In the future, very well educated expert teams will perform all four tasks in a more integrated and flexible way (indicated by the green line). From the technical point of view, a common microscopic database for all tasks will be the key issue. Microscopic data will increase the reliability of results and utilising the appropriate technology will result in rapid turn around times.

The use of clearly defined and permanently maintained interfaces for the data transfer will be essential. Probably only a few product oriented software tools will be in use in the future. These products will have to be very flexible. They will have a modular design that enables them to respond to different business demands. The advantage for RC and TOC then will be the possibility to reduce development costs significantly and to share the risks of any new development with several users.

As another result, any difference – if they ever exist – between “planning” and “production” tasks and systems will disappear because both tasks will amalgamate into each other. On the other hand, the complexity of all these processes will increase and will need excellent management.

6 Summary

The paper provides a short overview of the fundamental basics about which modelling techniques should be applied for the timetable design and the simulation of railway operation. The focus has been set on the microscopic



infrastructure models pointing out new techniques of flexible infrastructure modelling with a “Multi Scale” approach.

The introduction of the Seven-Step-Model illustrates major steps of modelling. This model transferred some feedback from several hundred projects performed by IVE, RMCon and RailSys customers over recent years into the modelling process. The next challenges will be to find a straightforward way to establish microscopic modelling as the basis for all planning purposes. The impact of such an approach may be that very well educated expert teams will lead not only one process but several.

References

- [1] <http://www.rmcon.de>
- [2] <http://www.ive.uni-hannover.de>
- [3] Siefer, Th., Radtke, A.: “Lectures Summer School Tongji University Shanghai, China 2005”.
- [4] Radtke, A., Hauptmann, D.: “Automated planning of timetables in large railway networks using a microscopic data basis and railway simulation techniques”, Comrail Dresden 2004, Proceedings page 615 - 624
- [5] Radtke, A.: “EDV-Verfahren zur Modellierung des Eisenbahnbetriebs”, Habilitation an der Universität Hannover 2005



This page intentionally left blank

The contribution of state resources in a constraint-based scheduling model for conflict solving at railway junctions

J. Rodriguez

INRETS (Institut National de Recherche sur les Transports et leur Sécurité), Villeneuve d'Ascq, France

Abstract

Scheduling is the process of allocating resources to activities over time. In a scheduling problem, resources are scarce and constrained in various ways (e.g. in the order of activities and the capacity of resources), and one is looking for a schedule of the activities that both satisfies the constraints and is optimal according to some criterion.

Over the last decade, constraint-based scheduling (CBS) has become the dominant form of modelling and solving scheduling problems.

This paper deals with a CBS model of real time management of train traffic through junctions, it focuses on the use of state resources to improve the model of conflicts between trains running in opposite directions.

1 Introduction

The railway industry has to improve the quality of service provided in order to increase freight and passenger market-shares. One important parameter that affects the quality of service is the efficiency and the effective use of resources. To achieve this aim, one solution is the use of computer-aided systems in planning and traffic control. A significant part of these computer-aided systems involves the model and the solution method of the optimisation problem, associated with real situations and decisions to be taken.

When traffic on connecting lines of a junction is heavy, the junction is likely to be a critical resource. As a consequence, small disruptions are amplified. A few seconds originally can quickly lead to a delay of more than 5 minutes. To limit this



phenomenon, new tracks, points and fly-overs can be built. Space and investments are needed for such extensions. However, such solutions are unlikely to be feasible in the short or medium term or impossible to implement in urban areas.

This provides the motivation for the study of new methods and models to optimise the use of scarce resources like rail junctions.

Concerning traffic control at a junction, the operator must perform the following tasks :

1. Analyse and select relevant information coming from the field,
2. Compare data with planned schedules,
3. Detect or anticipate conflicts,
4. Select and evaluate alternative solutions that reduce delays caused by conflicts,
5. Choose and implement a solution.

In this context, a computer-based assistance can be used for the first four tasks to improve the quality of the final solution. Task 4 can be formulated as an optimisation problem, the decision variables are the *selection of the alternative routes and sequences for trains* and the criterion is the *sum of delays*. This model and the solution method is part of the computer-aid system; nevertheless, the final decision must be left to the operator.

During the last decade, the problem of railway traffic management has been many times addressed by using a constraint-based scheduling (CBS) approach. An earlier attempt was in [1]: the application was the compilation of a railway timetable i.e. to construct a schedule of trains arrival-departure at each station of a line. The arrival-departure times must satisfy a set of temporal constraints.

Further to this contribution, constraint propagation and scheduling theory has been drawing the attention of researchers to solve railway traffic management problems. We can mention here some studies and applications :

- The generation of timetables for the suburban railway network of Melbourne [2],
- SAGITAIRE : a tool for the design of one day operation of the schedules, platform and route assignments of the trains for “Gare du Nord” station in Paris [3],
- PRaCoSy : a project for the automation of the preparation and updating of the running map for a section of the Beijing-Guangzhou line in China [4],
- CAPRES : a tool to analyse the capacity of railway networks with a saturation method [5],
- The scheduling of trains on single track networks [6, 7],
- COMBINE2 : a European project which aim was to define a conflict resolution model to use in control centres for fixed and moving block signalling systems [8],
- A generic tool for the National Network of Spanish Railways which covers many functions such as analysing the capacity, identifying bottlenecks, providing alternative planning and real-time traffic control [9].

Our first CBS formulation of the traffic management problem at junction was in [10], the last updated and detailed description of the model was in [11].



In this paper, we present an extension of the model of [11] by using *state resource constraints*. These kinds of constraint are redundant constraints. The state resource constraints allow a better management of conflicts between trains running in opposite directions in a section of track.

The paper is organised as follows: our CBS formulation of the railway traffic management is introduced in section 2. Within the framework of this model, we define in section 3 the opposite direction conflicts and the state resources associated. In section 4 the constraints between the activities and the state resources are defined according to different value assignments of the decision variables.

Finally, section 5 gives some preliminary results and issues.

2 Constraint-based scheduling model

The basic idea of the model is that *a train passing through a junction is a job*. According to scheduling theory, the concept of job is a set of activities linked by a set of precedence constraints. The movement of a train is a sequence of activities. Each activity is an elementary movement of the train through a track circuit. This is illustrated in figure 1.

As the train remains on track circuit until the next one becomes available for running, this limitation is named a “blocking constraint” in scheduling theory. Therefore our model is similar to that of *blocking job shop scheduling problem* [8, 12].

The constraints of our model will be roughly introduced, there is a more detailed formulation in [11]. The constraints of the problem can be formulated as follows:

- As each track circuit is a resource, the choice of a route for a train is turned into resource assignments for a sequence of activities. A constraint enumerates the combination of tuples of values allowed for the route and track circuit variables.
- The track circuits are modelled as unary resources, this leads to the constraint that two activities requiring the resource cannot overlap in time.
- Within the duration of an activity, we distinguish the detection phase. For each train, a constraint links the route values with the earliest start and finish time of detection phase of each activity.
- For each activity, a waiting time variable models the time spent when the next resource is not available. This time is added in the expression of the duration of the activity.
- The headway constraint between successive trains due to the block signalling system is formulated with a “synchronisation constraint”. Let us consider a block signalling system with two aspects. In that case, a train enters a block if no train is detected inside. Therefore, to enter a block, all track circuits inside the block must be available at the same time. The start of each activity related to one block has to be *synchronised* with the start of the detection on the first track circuit of the block. For the general case of a block system with n aspects, the synchronisation is established with



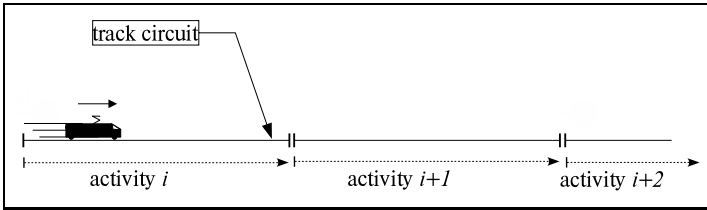


Figure 1: Train movement as a sequence of activities.

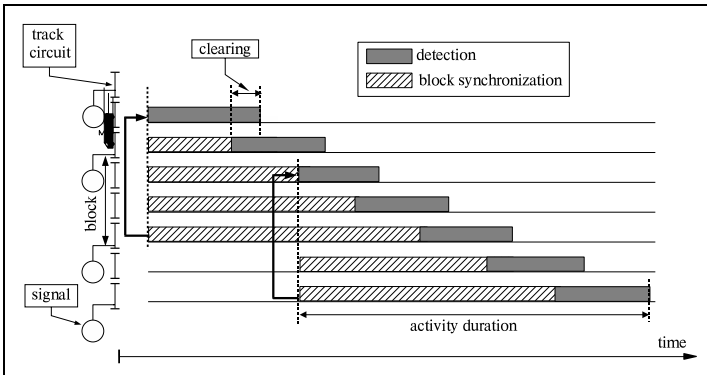


Figure 2: Gantt chart of activities for 3-aspect block signalling system.

the entrance in the first track circuit of the $n - 2^{th}$ previous block (e.g. see hatched rectangles for $n = 3$ in figure 2).

For train scheduling, the criterion frequently used is the sum of train delays caused by conflicts. This criterion is formulated with the sum of the waiting time variables.

3 Opposite direction conflicts

Let us define a “conflict sequence” as a common sequence of track circuits requested by two trains. There will be a running conflict when both trains request one track circuit of the conflict sequence during a common time interval. To arbitrate the conflict, the train circulations are ranked on the conflict sequence, i.e. the corresponding activities are ranked according to the train rank decision. For the case of a conflict sequence between routes running in the same direction, a rank decision on one activity implies all ranks of the other activities. This propagation of the rank decision is due to the blocking constraint (see section 2). However, in the case of an opposite direction conflict, there is no propagation of the rank decision; all the activities of the conflict sequence have to be ranked. The search algorithm for a feasible schedule is based on rank decision of activities. When there are opposite direction conflicts, the search algorithm will slow down as it needs more backtracks to find feasible ranks.



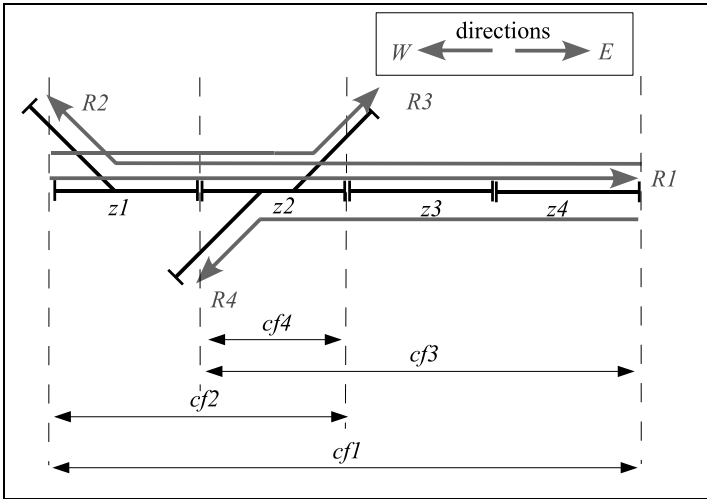


Figure 3: Example of 4 opposite conflict sequences.

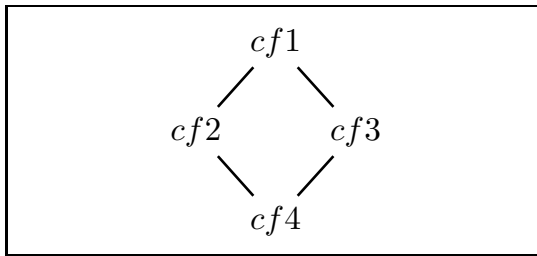


Figure 4: The Hasse diagram of conflict sequences.

To propagate the rank decision along the activities of an opposite direction conflict, we suggest using a state resource. A state resource is characterised as follows:

- It has an infinite capacity,
- The state can vary over time.
- Each activity may, during its execution, require it to be in a given state.
- Two activities may not overlap if they require incompatible states.

In the context of a railway traffic optimisation problem, a state resource models the allowed running direction during time along the conflict sequence. Therefore a state resource with 2 states (one for each direction) is associated with each opposite conflict sequence. The next section describes the constraints to post in order to obtain a correct propagation of rank decisions.



4 State resource constraints

The state resource constraints that can be posted depend on the resource assignment variables chosen for extending a partial solution during the search. Two cases have been considered: route assignment and track circuit assignment.

4.1 Route assignment

In this section we consider that the state resource constraints are posted after the routes are assigned to trains during a search algorithm. When a train is entering on the conflict sequence, the running direction will be constant from the start time of the first activities until the end time of the last activity. To satisfy this requirement, we define a constraint requiring that the state of the state resource has be constant during all the activities of the conflict sequence.

To formulate the constraint, let $Ct(t, cfi, state)$ be the constraint of the state resource of cfi in state $state$ for a train t . This constraint is the conjunction of the state resource constraints of all activities of the conflict sequence i.e. :

$$Ct(t, cfi, state) = \bigwedge_{z_j \in cfi} activity(t, z_j).requires(cfi.resource, state)$$

To illustrate this constraint, let us consider the example of figure 3. The opposite direction of the routes $R1$ and $R2$ yields the conflict sequence $cf1 : < z1, z2, z3, z4 >$. If a train is running on route $R1$, the state resource associated to $cf1$ must be on the state E during the running through all the track circuits of $cf1$. The state constraint of the route $R1$ is then :

$$\begin{aligned} Ct(t, cf1, E) = & activity(t, z1).requires(cf1.resource, E) \wedge \\ & activity(t, z2).requires(cf1.resource, E) \wedge \\ & activity(t, z3).requires(cf1.resource, E) \wedge \\ & activity(t, z4).requires(cf1.resource, E) . \end{aligned}$$

The state resource of the route $R2$ is similar, but with the state W .

The rank decisions to arbitrate a conflict between two trains t_1, t_2 requiring any track circuit of $cf1$ will be translated as one of the rank decisions :

- $activity(t_1, z4) \prec activity(t_2, z4)$,
- $activity(t_1, z1) \prec activity(t_2, z1)$.

This results is obtained by the constraints $Ct(t1, cf1, E)$ and $Ct(t2, cf1, W)$ which propagates the decision along the conflict sequence.

4.2 Track circuit assignment

Let now consider the case of extending a partial solution of a problem with track circuits assignments to trains. The question is then : what are the sufficient conditions to post a state resource constraint ? To answer the question, let us consider again the example of figure 3. There are 4 conflict sequences :

- $cf1 : < z1, z2, z3, z4 >$,



- $cf2 :< z1, z2 >$,
- $cf3 :< z2, z3, z4 >$,
- $cf4 :< z2 >$.

The conflict sequence set with the inclusion relation (symbol $\subseteq$) is a partially ordered set. This order can be graphically displayed as a Hasse diagram. Figure 4 is the Hasse diagram of the conflict sequences of the example of 3. In the Hasse diagram, each conflict sequence is represented as a vertex and a line is drawn upward from cf_i to cf_j if $cf_i \subseteq cf_j$, and there is no cf_k such that $cf_i \subseteq cf_k \subseteq cf_j$, we say that cf_j covers cf_i .

When a track circuit is assigned to a train, the following rules are used to post constraints:

Rule 1: If a the track circuit z_i is assigned to a train then post the state resource constraint of *least conflict sequence* which includes z_i (noted $lcs(z_i)$).

Rule 2: For all conflict sequence cf_i covered by $lcs(z_i)$, post state resource constraint of cf_i .

To illustrate Rule 1, let us consider the example of figure 3, and the case of the assignment of track circuit $z1$ to a train in the E direction. According to the rule, the state resource constraint requiring the conflict sequence $cf2$ in E direction is posted. Indeed, $cf2$ is the $lcs(z1)$, i.e. none of the conflict sequences which include $z1$ is a subset of $cf2$. Note that $cf1$ includes $z1$, but it is not the $lcs(z1)$ as $cf2$ is one of the subsets of $cf1$ (see the upward line of the Hasse diagram in figure 4).

The second rule concerns the conflict sequence covered by the least conflict sequence. We can take the as an example the figure 3 and the case of the assignment of track circuit $z1$. As the resource of $cf2$ is required by the train in direction E, to avoid inconsistencies, we must require the resource of the covered conflict sequence $cf4$ in the same direction and during the same interval.

So, after the assignment of $z1$ to the train, the state resources of $cf2$ and $cf4$ will be required in the state E, whatever the route assigned afterwards.

5 Results and conclusion

The extension of our CBS model presented here has been implemented and successfully validated on the infrastructure of the Pierrefitte-Gonesse junction. Table 1 shows some preliminary results on a set of growing size problem instances.

The column headings in the table have the following meanings:

#t	: Number of trains of the problem instance,
CBS	: CBS model of section 2 ,
CBS+SRC	: CBS model with the state resource constraints extension,
Crit	: Criteria of the best solution found,
CPU	: Computation time in CPU seconds to find the best solution,
BT	: Number of backtracks of the search algorithm.



Table 1: Computational results.

Inst.	#t	CBS			CBS+SRC		
		Crit	CPU ^a	BT	Crit	CPU ^a	BT
1	6	64	1.63	206	64	1.91	138
2	8	64	1.75	631	64	2.42	322
3	8	66	2.4	721	66	3.1	304
4	10	161	33.42	8774	161	22.33	1172
5	12	161	68.66	9848	161	43.91	1286
6	14	313	339.6	27 323	258	832.12	13 646
7	16	350	704.93	24 789	258	1199.3	14 379
8	18	498	955.67	9055	461	646.02	7739
9	20	512	345.2	11 994	498	1585.14	7611
10	22	648	718.08	5084	609	1593.16	4135

^aIn seconds using a 2.66GHz Pentium IV processor, Ilog Solver 5.3 and Ilog Scheduler 5.2

The bold face values show the improvements of the state resource constraints extension of the CBS model.

For the problem instances used here, searching for an optimum value is not a realistic goal. In order to apply a resolution method and obtain *good solutions within a limited time*, an heuristic approach has been considered (see [11]). The two models have been tested with the same CPU limits.

The first observation is that the number of backtracks has been significantly reduced for all instances. As we also obtain a better or the same criteria solution, this shows that the redundant constraints have pruned many branches of the search tree. For example in the instance #5, the number of backtracks has been reduced from 9848 to 1286 and the criteria has been reduced from 68.66 to 43.91.

In what regards the criteria values, we remark that the improvements have been obtained with the large instances. This can be explained by the computational time needed to calculate the state resource constraints conditions (see section 4). This time increases very slowly with the instance sizes and therefore has more impact in small instances than in large instances.

As a conclusion, the state resource constraints show very promising preliminary results in what regards the resolution performances. This first attempt of using these kinds of constraint need further researches. Some improvement issues have been identified regarding :

- The reduction of the number of constraints needed,
- The combination of assignment and scheduling heuristics of the solution algorithm,
- The testing of the practical limits of some complete scheduling algorithms,



- An interesting future task is also to study in depth the performances of different models and resolution methods as initiated in [13], having in mind the goal of the potential integrations of different approaches.

References

- [1] Fukumori, K. & Sano, H., Fundamental algorithm for train scheduling based on artificial intelligence. *Systems and Computers in Japan*, **18(3)**, 1987.
- [2] Gosselin, V., Train scheduling using constraint programming techniques. *Charme Technical Bulletin*, (3), 1993.
- [3] Poujade, M., Moyon, P. & J.F., P., Sagitaire: Allocation des quais de la gare Paris-Nord. *Conférence IA'95*, 1995. (in french).
- [4] Chiu, C.K., Chou, C.M., Lee, J.H.M., Leung, H.F. & Leung, Y.W., A constraint-based interactive train rescheduling tool. *Second International Conference on Principles and Practice of Constraint Programming*, pp. 104–118, 1996.
- [5] Hachemane, P., *Évaluation de la capacité de réseaux ferroviaires*. Thèse 1632, École Polytechnique Fédérale de Lausanne, 1997. (in french).
- [6] Kreuger, P., Carlson, M., Olsson, J., Sj, T. & o Astr, Trips scheduling on single track networks - the tuff train scheduler. *CP'97 Workshop on Industrial Constraint - Directed Scheduling*, pp. 1–12, 1997.
- [7] Oliveira, E.S., *Solving Single-Track Railway Scheduling Problem Using Constraint Programming*. Ph.D. thesis, School of Computing - University of Leeds, 2001.
- [8] COMBINE2 project IST-2001-34705, Traffic regulation and co-operation methodologies. Deliverable D3, Univerita degli Studi Roma Tre, 2003.
- [9] Barber, F., Salido, M.A., Abril, M., Ingolotti, L., Lova, A. & Tormos, P., An interactive train scheduling tool for solving and plotting running maps. *LNCS/LNAI*, (3040), pp. 659–668, 2004.
- [10] Rodriguez, J. & Kermad, L., Constraint programming for real-time train circulation management problem in railway nodes. *Computers in Railways VI*, WITPress: Lisbonne-Portugal, 1998.
- [11] Rodriguez, J., A constraint programming model for real-time trains scheduling at junctions. *Transportation Research Part B : Methodological*, 2006. In press.
- [12] Mascis, A. & Pacciarelli, D., Job-shop with blocking and no-wait constraints. *European Journal of Operational Research*, (143), pp. 498–517, 2002.
- [13] Delorme, X., Rodriguez, J. & Gandibleux, X., Heuristics for railway infrastructure saturation. *Electronic Notes in Theoretical Computer Science 50 NI(2001)* URL: <http://www.elsevier.nl/locate/entcs/volume50.html>, pp. 41–55, 2001.



This page intentionally left blank

A new idea for train scheduling using ant colony optimization

K. Ghoseiri

*School of Railway Engineering,
Iran University of Science and Technology, Iran*

Abstract

This paper develops a new algorithm to the train scheduling problem using Ant Colony System (ACS) meta-heuristic. At first, a mathematical model for a kind of train scheduling problem is developed and then the algorithm based on the meta-heuristic is presented to solve the problem. The problem is considered as a traveling salesman problem (TSP) wherein cities represent the trains. ACS determines the sequence of trains dispatched on the graph of the TSP. Based on the sequence obtained and removing for collisions incurred, train scheduling is determined. Numerical examples in small and medium sizes are solved by using the algorithm. The solutions are compared to the exact optimum solutions to check for quality and accuracy. Comparison of the solutions shows that the algorithm results in good enough quality and time savings. A case study is presented to illustrate the solution.

Keywords: meta-heuristic, ant colony optimization, Ant Colony System, train scheduling problem, traveling salesman problem.

1 Introduction

Rail transport planning is a very complex task which is carried out based on the mutual reaction among a large number of impressed components. As it was mentioned in Ghoseiri et al [5] and Lindner [10], in respect of complexity of rail transport planning, this process is divided into several steps. These procedures include the demand analysis, line planning, train scheduling, rolling stock planning and crew management [5]. These problems are NP-hard problems and solving them with exact algorithms is very difficult and time consuming.



Although metaheuristic algorithms produce the solutions close to optimum, there is a little research in the rail transport planning problems. As an example, Huntley et al [8] developed a simulated annealing approach to train scheduling for CSX Transportation. Van Wezel et al [15] applied Genetic Algorithm to improve the timetable. Martinelli and Teng [11] used the Neural Network to routing in a railway. Nachtigall and Voget [12] applied the Genetic Algorithm to solve train scheduling problems. Gorman [6] used a combination of Genetic algorithm and Tabu Search for addressing the weekly routing and scheduling problem. Pacciarelli and Pranzo [14] used the Tabu Search to solve train scheduling problem. Kwan and Mistry [9] used a Co-evolutionary Algorithm to create a train timetable. Engelhardt-Funke and Kolonko [3] used an advanced evolutionary algorithm to solve train scheduling problems.

Fischetti et al [4], Gutin and Punnen [7] and Noon and Bean [13] showed that the train scheduling problem can be transformed to a travel salesman problem easily. Dorigo and Gambardella [2] showed that ACS algorithm has been more successful than the other metaheuristics in solving TSP. Therefore, considering the capability of transforming train scheduling problem to a TSP problem, good responses can be expected from solving it by ACS algorithm. In solving the train scheduling problems, ACO algorithms are not used so far.

In this article, by considering the good quality of ACS algorithm in solving TSP problem and also transforming capability of train scheduling problem to TSP it is decided to solve the train scheduling problem by this algorithm.

2 ACS compound model and train scheduling problem

Train scheduling is a combinatorial optimization problem. In this problem the aim is to determine the arrival and departure times from stations which train passes on them. This problem is known to be NP- hard type. Because of dimension and natural complexity in mathematical models, traditional optimization techniques are not useful for solving the problem and the accurate methods are just usable with examples in small sizes and for solving the problem with real dimension, the heuristic methods should be used. In this article, for solving the train scheduling problem ACS algorithm is chosen.

2.1 Ant colony system (ACS)

ACS was suggested as a new heuristic method to solve optimization problem by Dorigo and Gambardella [1]. This is the reformed form of AS algorithm and functions as follow:

Each ant generates a complete solution by choosing the nodes according to a probabilistic state transition rule.

$$s = \begin{cases} \arg [\max_{j \in N_i^k} \{[\tau_{ij}][\eta_{ij}]^\beta\}] & \text{if } q \leq q_0 \\ S & \text{if } q > q_0 \end{cases} \quad (1)$$



$$p_{ij}^k = \frac{[\tau_{ij}][\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{ij}][\eta_{ij}]^\beta} \quad (2)$$

Where q is a random number uniformly distributed in $[0 \dots 1]$, q_0 is a parameter between 0 and 1 and S is a random variable selected according to the probability distribution given in Eq. (2), τ_{ij} is the amount of pheromone in edge ij ,

$\eta_{ij} = \frac{1}{\delta_{ij}}$ that δ_{ij} is the cost of edge ij , β determine the relative importance

of η versus τ and N_i^k is the remaining nodes set of ant k based on moving from node i to build an feasible solution.

The state transition rule given in Eq. (1) and Eq. (2) is called pseudo-random-proportional rule. In ACS, only globally best ant which has built the best solution, deposits pheromone in the graph. When all the ants built their solutions, global updating rule of pheromone is applied. This rule is as follow:

$$\tau_{ij} \leftarrow (1 - \rho)\tau_{ij} + \rho\Delta\tau_{ij} \quad (3)$$

where $0 < \rho < 1$ is pheromone decay parameter and $\Delta\tau_{ij}$ equals to:

$$\Delta\tau_{ij} = \begin{cases} \frac{1}{\text{cost}_{gb}} & \text{if } (i,j) \in \psi^{gb} \\ 0 & \text{if } (i,j) \notin \psi^{gb} \end{cases} \quad (4)$$

ψ^{gb} is the best solution which was built and cost_{gb} is the cost of the best solution. While building a solution by using local updating rule of pheromone in this method, the pheromone level is reduced by each ant in visiting edges.

$$\tau_{ij} \leftarrow (1 - \xi)\tau_{ij} + \xi\tau_0 \quad (5)$$

τ_0 is a small fixed value and $0 < \xi < 1$ is the local evaporation coefficient of pheromone.

2.2 The proposed mathematical model of train scheduling

This model can be used for a single track and double track without any branches. In this model it is supposed that the trains are just dispatched from the first and last station. After preparation, the trains in the beginning or end stations should be dispatched immediately. In the case that the prepared trains to dispatch are stopped in the stations with unpermitted time stop, we undergo some cost. In this model, the speed and trip times in each track section for each train are supposed as a fixed value. In this model the train can travel in two directions, but they are not permitted to overtake.

2.2.1 Notation

R: the group of trains that should be dispatched from right station to left.

L: the group of trains that should be dispatched from left station to right.



T: the group of total trains. ($i, j \in R$ or L or T and $T = R \cup L$)

B: set of track sections. ($k \in B$)

Track section k is a section of track that connects two stations k and $k+1$.

Track sections and stations are indexed in numerical order from left to right.

S: set of stations. ($k \in S$)

D: the set of permitted stop time in the station. ($d_{ij} \in D$)

AD: the set of arrival and departure times from station. ($Xa(i, k), Xd(i, k) \in AD$)

2.2.2 Parameters

Trip time: the time that train i needs to pass track section k . (t_{ik})

Dwell time: this time indicates the permitted dwell time of train i in station k . (d_{ij})

Headway: minimum time interval between trains i and j to arrive/depart in/from track section k . (h_{ijk})

Train importance weight: (W_i)

2.2.3 Decision making variables

Binary variables:

$$a_{ij} = \begin{cases} 1 & \text{if train } j \in R \text{ enter the track section after train } i \in R \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$b_{ij} = \begin{cases} 1 & \text{if train } j \in L \text{ enter the track section after train } i \in L \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$c_{ijk} = \begin{cases} 1 & \text{if train } j \in L \text{ enter the track section } k \text{ after train } i \in R \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Continuous variables:

$Xa(i, k)$: the arrival time of train i to station k .

$Xd(i, k)$: the departure time of train i from station k .

2.2.4 Objective function

Objective function in this model is to minimize the total train delays in the stations. The delay equals time difference of train stop time in a station with the permitted dwell time in the station.

$$\text{Min } \mathbf{z} = \sum_{i \in T} \sum_{k \in S} W_i (Xd(i, k) - Xa(i, k) - d_{ik}) \quad (9)$$

2.2.5 Constraints

Trip times constraints of dispatched trains from right station:

$$Xa(i, k) - Xd(i, k-1) = t_{ik} ; i \in R, k \in S \quad (10)$$

Trip times constraints of dispatched trains from left station:

$$Xa(i, k-1) - Xd(i, k) = t_{ik} ; i \in L, k \in S \quad (11)$$



Stop times constraints of dispatched trains from left and right stations:

$$Xd(i, k) - Xa(i, k) \geq d_{ik} ; i \in T, k \in S \quad (12)$$

Sequence constraints of dispatched trains from right station:

$$\begin{aligned} Xd(j, k-1) - Xa(i, k) &\geq h_{jk} - M(1 - a_{ij}) \\ Xd(i, k-1) - Xa(j, k) &\geq h_{jk} - Ma_{ij} \end{aligned} \quad (13)$$

$$i, j \in R, k \in S$$

Sequence constraints of dispatched trains from left station:

$$\begin{aligned} Xd(j, k) - Xa(i, k-1) &\geq h_{jk} - M(1 - b_{ij}) \\ Xd(i, k) - Xa(j, k-1) &\geq h_{jk} - Mb_{ij} \end{aligned} \quad (14)$$

$$i, j \in L, k \in S$$

Safety constraints that ensure no collision occur between two trains of opposite directions:

$$\begin{aligned} Xd(i, k) - Xa(j, k) &\geq h_{ijk} - Mc_{ijk} \\ Xd(j, k+1) - Xa(i, k+1) &\geq h_{ijk} - M(1 - c_{ijk}) \end{aligned} \quad (15)$$

$$i \in R, j \in L, k \in S$$

2.3 The solution method of the proposed model using ACS

In the proposed algorithm, it is supposed that the trains play the role of cities (nodes) in TSP. The dispatched trains from left to right and also dispatched trains from right to left form two independent sub-networks in TSP. According to the definition, selected path of each ant in the trains' network indicates the sequence of train dispatching.

For instance, in figure 1 which includes 7 trains (3 dispatching trains from right to left and 4 dispatching trains from left to right), if an ant chooses the path of Start node, train 1, train 3 and train 2 it mean the dispatching sequence of trains 1,3,2.

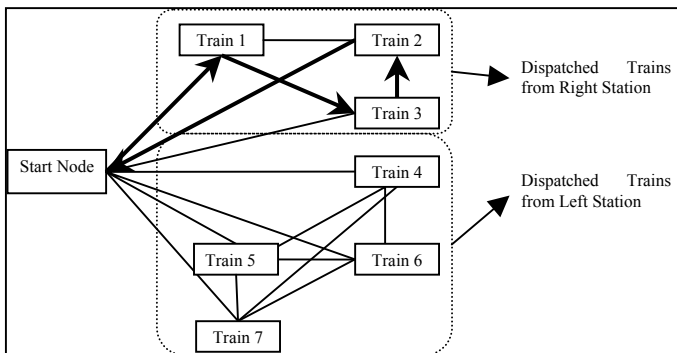


Figure 1: Problem's graph for the example of seven trains.

In this algorithm, a colony induces $2 \times n$ ants. Each group includes 2 ants that one of them is allocated to build the sequence in set of dispatched trains from right to left station and another one is allocated to build sequence in dispatched trains from left to right station.

At first both ants are placed in figurative node of zero (the start node). Then one of the ants is chosen from the first group randomly. The first chosen ant chooses a train in its train group by using pseudo-random-proportional rule. The arrival and departure times of this train from start station to the final station is calculated. Then another ant chooses its train. The arrival departure of this train from each station is identified in regard to collision with the opposite train. If the chosen train is the dispatched train from the left station, in the case that the obtained times are true in equation (16), the collision occurs:

$$\begin{aligned} Xa(i, k) + \text{headway} &> Xd(j, k) \\ &\& \\ Xd(i, K - 1) - \text{headway} &< Xd(j, k - 1) \end{aligned} \quad (16)$$

In this equation j is the selected train, i indicates chosen trains from the group of dispatched trains from right station and k is a track section in which the collision occurred. In this case for collision resolution between two trains, the departure time of chosen train from the related station is changed as follows:

$$Xd(j, k) = Xa(i, k) + \text{headway} \quad (17)$$

The arrival and departure times of this train to its last station is calculated based on this time.

If the chosen train is the dispatched train from right station, in this case the obtained times are true in equation (18), therefore, the collision is occurred.

$$\begin{aligned} Xa(i, k - 1) + \text{headway} &> Xd(j, k - 1) \\ &\& \\ Xd(i, k) - \text{headway} &< Xd(j, k) \end{aligned} \quad (18)$$

In this equation, j is the chosen train, i indicates the chosen trains from the group of dispatched trains from left station and k is a track section in which the collision is occurred. In this case for resolution of collision between two trains, departure time of above chosen train from related station changes as follows:

$$Xd(j, k - 1) = Xa(i, k - 1) + \text{headway} \quad (19)$$

The chosen trains are omitted from the set of trains. Then randomly an ant is selected again. This ant chooses a train from its group. The arrival and departure times of this train is identified with its chosen sequence in its group. When the arrival and departure time from a section was identified, the collision condition of chosen train with dispatched chosen train in opposite direction is checked. In the case of collision, it is removed. This operation continues in the same way so that all the arrival and departure times from all stations are identified and there are not any collisions in the sections. Then the next train is chosen by other ants. This procedure continues until ants choose the all trains of their own group.



3 The result analysis of the model

To analysis the solution results obtained from ACS-TS are compared with those of exact optimization method of the train scheduling model. For this purpose, computations are carried out for 28 problems. Figure 2 shows the run times of solving the problems using LINGO software in comparison with the proposed ACS algorithm.

It was considerable that in 28 solved problems the overall delay amount in dispatching trains from both methods was almost equal. Although proposed ACS time solving method has shown considerable time saving in comparison to the exact solving method.

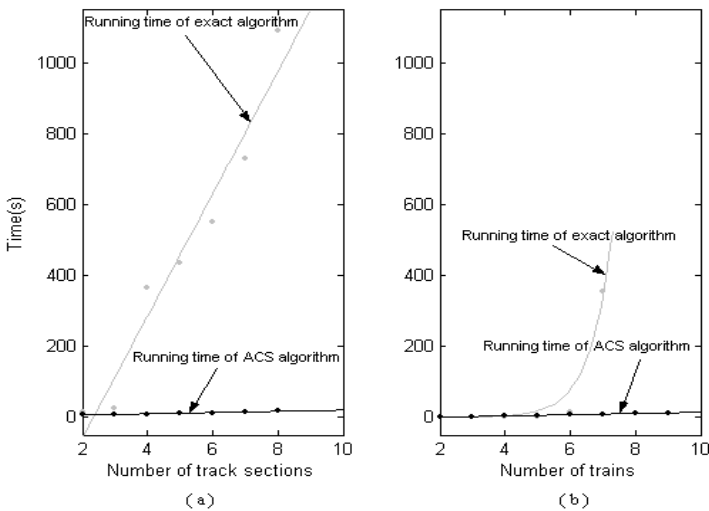


Figure 2: Run times of solving the train scheduling problems.

4 The case study

In this section to clarify the proposed algorithm much more, a problem with 30 trains and 4 sections is solved.

At first, according to Dorigo and Gambardella [1] the primary values for parameters are set to the following values:

$\rho = 0.1$, $\xi = 0.1$, Phomone primary amount in edges = 0.000005, and $q_0 = 0.9$.

Also according to definition of problem, the number of ants in this colony is considered as equal to 30 and the fix primary amount $\tau_0 = 0.012$ that is obtained by $\tau_0 = \frac{1}{(n.L_{nn})}$ where n is the number of trains and L_{nn} is the solution cost



produced by a heuristic method (For further study refer to Dorigo and Gambardella [1]. Furthermore by considering this fact that in proposed algorithm, the length (cost) of arcs doesn't have a meaning therefore by supposing $\beta = 0$, the length effect of edges is omitted in ACS.

Then the parameters best values were experimentally determined. For this purpose the parameter values changed and then algorithm was run ten times. After this step the best value was chosen and then the problem was solved with these parameters.

$$q_0 = 0.9, \rho = 0.35, \xi = 0.2, \tau_0 = 0.$$

After determining parameters, the proposed algorithm for a problem with 30 trains was run. At the end, the amount of delay in this state was equal to 2492.1. Figure 3 is the indicator of convergence in improving the solutions in each cycle of running the algorithm.

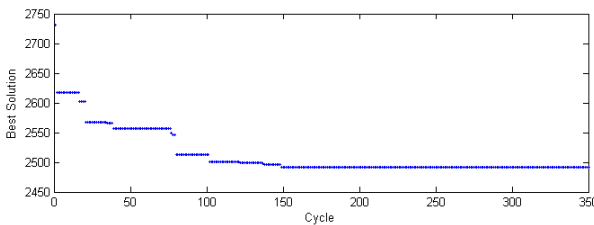


Figure 3: Convergence indicator in improving the solutions.

5 Conclusion

In this article, an algorithm was introduced for train scheduling problem by using ant colony optimization. At first a mathematical model was developed for a type of train scheduling problem and then for solving this model an algorithm was suggested based on ACS algorithm. According to the algorithm, the model transferred into a constructive graph in which each train was considered as a city in TSP graph. ACS algorithm identified the sequence of trains traveling in each direction on this graph and then based on this sequence the collision of the trains is removed and at the end the time-distance graph of trains movement was identified (either arrival or departure times). To check for quality of the solutions, numerical examples in small and average sizes were solved. The obtained results were compared with their exact optimum solutions. By comparing the results of exact solving of problems and solving them by the proposed algorithm, the time saving and good solution quality were concluded. At the end, a case study was presented for describing the computation manner.

References

- [1] Dorigo, M., & Gambardella, L.M., Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem. *IEEE Transactions on Evolutionary Computation*, 1(1), pp. 53–66, 1997.



- [2] Dorigo, M., & Gambardella, L.M., Ant colonies for the traveling salesman problem. *BioSystems*, 43, pp. 73–81, 1997.
- [3] Engelhardt-Funke, O., & Kolonko, M., Analysing stability and investments in railway networks using advanced evolutionary algorithms. *International Transactions in Operational Research*, 11, pp. 381–394, 2004.
- [4] Fischetti, M., Salazar-Gonzales, J., & Toth, P, The Generalized Traveling Salesman Problem and Orienteering Problem. *Traveling Salesman Problem and its Variations*, ed. G. Gutin and A. P. Punnen, Kluwer Academic Publishers, pp. 609–663, 2002.
- [5] Ghoseiri, K., Szidarovsky, F., & Asgharpour, M.J., A Multi-Objective Train Scheduling: Model and Solution. *Transportation Research part B*, 38, pp. 927–952, 2004.
- [6] Gorman, M.F., An Application of Genetic and Tabu Searches to the Freight Railroad Operating Plan Problem. *Annals of Operations Research*, 78, pp. 51–69, 1998.
- [7] Gutin, G., & Punnen A. P., *The Traveling Salesman Problem and its Variations*, Kluwer Academic Publishers, 2003.
- [8] Huntley, C.L., Brown, D.E., Sappington, D.E., & Markowicz, B.P., Freight Routing and Scheduling at CSX. *Transportation Interfaces*, 25(3), pp. 58–71, 1995.
- [9] Kwan, R.S.K., & Mistry, P., *A Co-evolutionary Algorithm for Train Timetabling*, Research Report Series No. 2003.13, University of Leeds, School of Computing, 2003.
- [10] Lindner, T., *Train Schedule Optimization in Public Rail Transport*, Ph.D. Thesis, Technische Universitat Braunschweig, 2000.
- [11] Martinelli, R.D., & Teng, H., Optimization of Railway Operations Using Neural Networks. *Transportation Research part C*, 4, pp. 33–49, 1996.
- [12] Nachtigall, K., & Voget, S., A Genetic Algorithm Approach to Periodic Railway. *Synchronization Computers and Operations Research*, 23, pp. 453–463, 1996.
- [13] Noon, C., & Bean, J., A Lagrangian Based Approach for The Asymmetric Generalized Traveling Salesman Problem. *Operations Research*, 39, pp. 623–632, 1991.
- [14] Pacciarelli, D., & Pranzo, M., A Tabu Search algorithm for the railway scheduling problem. *Proc. of the 4th Metaheuristic International Conference*, Porto (Portugal), pp. 16–20, 2001.
- [15] Van Wezel, M.C., Kok, J.N., Van den Berg, J., & Van Kampen, W., Genetic improvement of railway timetables. *Computer Science*, 866, pp. 566–574, 1994.



This page intentionally left blank

Joint design standard for running times, dwell times and headway times

V. A. Weeda & P. B. L. Wiggeraad

*Department of Transport & Planning,
Delft University of Technology, Netherlands*

Abstract

A punctual train service requires a feasible timetable. This paper deals with design standards for running times, dwell times, and headway times. Feedback from operations to planning is an appropriate way to determine the required standards for delivering the desired performance. This feedback has been explored by analysing the traffic process in recent months on two Dutch mainlines. A conceptual conflict model has been developed to derive an empirical relationship between headways and reactionary delays. The operational observations make up the basis for a flexible joint standard for running, dwell and headway times. The suggested standard explicitly takes conflicts into account, either preventing them by scheduling a larger headway time, or compensating for them by increasing the running time supplement of the succeeding train.

Keywords: conflicts, punctuality, running time, timetable design.

1 Introduction

1.1 Relevance

Punctuality is an important performance indicator in railway transport. A punctual railway service requires both an adequate planning and operation close to the planning [1]. This paper deals with the first of these and focuses on three elementary processes in railway operations: running times, dwell times, and headway times. The timetable reserves a particular time laps for each process, consisting of two parts: the technically minimal time, and a margin to cope with variations during operations. A certain quantity of margins is necessary for a



robust timetable; whether larger margins increase or decrease predictability is open to discussion. Clear is that too long margins raise production costs and, moreover, generate a less attractive railway service due to longer journey times or a limited train frequency. Well-considered allocation of margins could be done location-specifically, completely based on feedback from historical data. Although this method is very likely to produce a feasible timetable, it is quite an elaborate process that does not work in new situations. Therefore, this paper allocates margins by means of design standards. Determining the appropriate level of margins in these standards requires insight in two relationships:

- What is the quantitative relationship between margins and punctuality?
- As margins in running times, dwell times and headway times are all supposed to have a positive effect on punctuality, they should be interchangeable; according to which relationship?

1.2 Research method

These relationships have been investigated by statistical analysis of train running data and by analytical estimation of train delays. The railway traffic has been analysed in two case studies:

- Utrecht–Den Bosch, an electrified 48 km double-track mainline in the central part of the Netherlands, data from March 2004.
- Rotterdam–Dordrecht, an electrified 20 km four-track mainline in the West of the Netherlands, data from working days in November 2005.

These railway lines accommodate both inter-city services and local trains. Freight traffic runs on both lines as well, but this is not considered in the study.

Operational data are based on so-called TNV-logfiles: these files contain block section occupation and release data and other infra element status recordings from the Train Number Tracing system. The tool TNV-Prepare [2] extracts train runs and signal clearance times from these files. Recently, the logfiles have been enhanced with infrastructure data (the new recordings are called VTL-files). This format enables new functionalities for which the VTL-tool is developed (at NS Reizigers, the main Dutch passenger train operator). From measuring points at the front and end of a platform it calculates accurate arrival and departure times on the basis of actual rolling stock characteristics (train length, acceleration). It also determines unplanned stops in front of signals at danger.

Resulting train operation data have been broken down by type of rolling stock and by time of day. This decomposition yielded some minor differences but these do not influence the findings given in this paper and are not issued there.

1.3 Paper outline

The paper is structured as follows. Sections 2, 3 and 4 deal with running times, dwell times and headway times, respectively. These sections start with a short description of the current design standard; consequently, results of the operational analysis are presented. Section 5 gives an integrated view on the three types of processes and suggests a joint standard. Finally, section 6 draws conclusions.



2 Running times

2.1 Current standard

In current Dutch planning practice running times are calculated microscopically on the basis of infrastructure and rolling stock characteristics. On top of that, the planning standard prescribes a 7% supplement. This is supposed to cover daily variations in running times due to driving behaviour, weather circumstances, train load, etc. This standard is a minimum; capacity constraints and network coherence may force *planned* running times to be longer.

2.2 Operational practice

The feasibility of running times planned according to the standard has been assessed on the basis of autonomous runs, i.e. trains not held up by restrictive signals. Delays as a result of congestion effects are addressed separately in section 4. With help of signal status information in TNV data, perturbed runs can be identified and filtered. This filtering is reasonably accurate, but some cases of reactionary delays are still in the dataset, including unplanned stops at (not recorded) permissive signals.

The Utrecht–Den Bosch case study showed that inter-city trains make their runs within schedule, whereas local trains did not manage to do so. The next question was if the unfeasibility of local trains' paths was caused by running times, dwell times, or both. The relevant data, however, was too rough to isolate dwellings from running times. The Rotterdam–Dordrecht case study made this distinction possible. Apart from freight and international traffic, this line accommodates per hour per direction 4 inter-city services (running time approx. 14 minutes) and 6 local services with intermediate stops (running time 18–24 minutes).

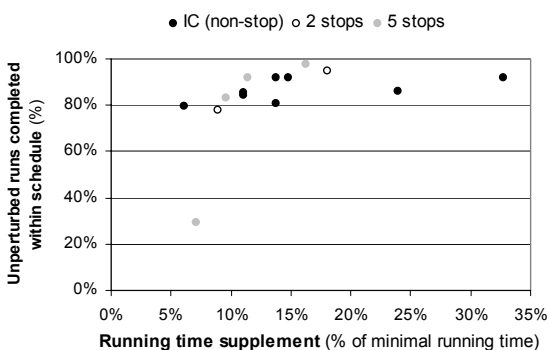


Figure 1: Feasibility of planned running times for autonomously running trains per train series on Rotterdam–Dordrecht.



Figure 1: displays for each train series the percentage of runs completed within scheduled running time, depending on the available running time supplement. Like between Utrecht and Den Bosch, a 7% supplement is already sufficient for about 80% of the unperturbed inter-city trains; a (much) larger supplement does not improve performance. Demanding a 100% running time feasibility is not necessary, taking into account that the dataset may still include conflicting runs or occasionally delays due to special causes along the way.

For local trains, the picture is slightly different. Supplements over 10% produce a feasible running time, but a path with only 7% appears to be too tight and can only be adhered to by 30% of the trains of the series. Given the data volume and accuracy, this is definitely not an outlier. To find an explanation why a 7% supplement is enough for inter-city trains but not for local trains, the next paragraph looks into the calculation of supplements.

2.3 Basis for running time supplements

Traditionally, the length of running time supplements is determined as a percentage of the minimal running time. This is not obvious, however. It is essential to realise that factors such as weather, train load and driving behaviour generally influence acceleration rather than top speed. The case study data shows that variations in running times are high in the accelerating and braking areas and much lower in the cruising area in-between. Hence, the need for supplements especially exists around stops rather than equally distributed along the way. Remarkably, local trains, stopping more frequently than inter-city trains, have no extra supplement but 7% of the running time. To this end, it seems logical to base the supplement on the number of stops instead of the minimal running time. This idea is worked out in section 5.

3 Dwell times

3.1 Current standard

Technically minimal dwell times are not as easy to calculate as minimal running or headway times. In a way, the minimal dwell time is just the time required to open and close the doors. Practically, the dwell time largely depends on the (widely varying) number of boarding and alighting passengers. Dwell times are therefore scheduled in different ways depending on station importance:

- Depending on passenger numbers main stations have a minimal dwell time of 1, 2 or 3 minutes for inter-city trains. In the same station categories the standards for local trains are 1, 1 and 2 minutes, respectively.
- At line-side stops, the minimal dwell time varies from 0.5 to 0.7 minute, determined by the type of rolling stock (depending on door width and spacing, amongst other factors). In the near future, a 1 minute dwell time will be required at a couple of busy line-side stops. As a rule, departure times are rounded down to avoid waiting and wasting time.

In planning dwell times, the concept of margins is not used explicitly.



3.2 Operational practice

Just as the running time analysis, the dwell time analysis focuses on Rotterdam–Dordrecht. The dwell time analysis analogously excluded stops where the train had to wait for the departure signal to turn clear. Early arriving trains have been removed from datasets as well, as they have to await scheduled departure and thus are not representative for minimal dwelling.

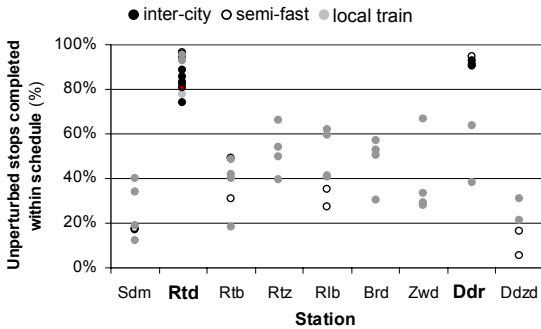


Figure 2: Feasibility of planned dwell times for autonomous stops per train series on Rotterdam–Dordrecht.

Figure 2: features the percentage of trains in a series that meets the scheduled dwell time per station. Stops exceeding the standard are left out. The remaining stops are scheduled at 2 minutes at Rotterdam (Rtd) and Dordrecht (Ddr), but 1 minute for local trains at Ddr, and 0.6 or 0.7 minute at line-side stops. The picture is clear:

- Dwell times at main stations are feasible for all inter-city trains. The 1 minute standard for local trains (Ddr) is questionable: the operational average exceeds the standard.
- Dwell times at line-side stops are unrealistic. At some spots they are even very unrealistic, only being met by about 30% of the trains.

4 Headway times

4.1 Current standard

In this project, the term “headway” applies to any situation where two trains consequently use the same infrastructure (track, switch or crossing). In contrast with running times, headway times are in the Netherlands planned in a macroscopic way. Generally, the minimal interval between two train movements at the same location is 3 minutes, although other (integer) headway times hold for certain types of successions and for certain locations. The difference between the technically minimal headway time and the planned interval is available as buffer time, allowing a train to run slightly late without consequences for other trains.



4.2 Conceptual model infrastructure conflict

If a train's delay exceeds the buffer time, it may cause reactionary delays to succeeding trains. This mechanism has been studied with a newly developed conceptual model, which will be enlightened with an example from the Utrecht–Den Bosch case study.

The situation is a level junction where two consecutively train movements ("train 1" and "train 2") have a crossing in common. Their scheduled headway time is 3 minutes with 44 seconds of buffer time. Before the trains meet and possibly influence each other, both of them have a certain deviation from the timetable (positive, zero or negative). These deviations are derived from the TNV data and plotted against one another in Figure 3. Each of these approx. 700 dots represents one couple of train 1 and train 2. For every dot, it is determined whether train 1 held up train 2. A few conditions divide the dots into a few categories:

- Ia Train 1 is running on time or has a delay smaller than the buffer time. Train 2 is on time or late and is not perturbed (40%).
- Ib Train 2 is running early and is held up (under the diagonal dotted line). It does not matter since it only makes train 2 run on time again (16%).
- Ic Train 1 is late but train 2 is even later and does not suffer an additional delay. Although train 2 should not run late, there is no headway conflict (12%).
- II Train 2 reaches the junction before it is released by train 1. A conflict occurs: train 1 delays train 2 or occasionally the other way round (23%).
- III Train 1 is running that late, that the train dispatcher gave priority to train 2 (9%).

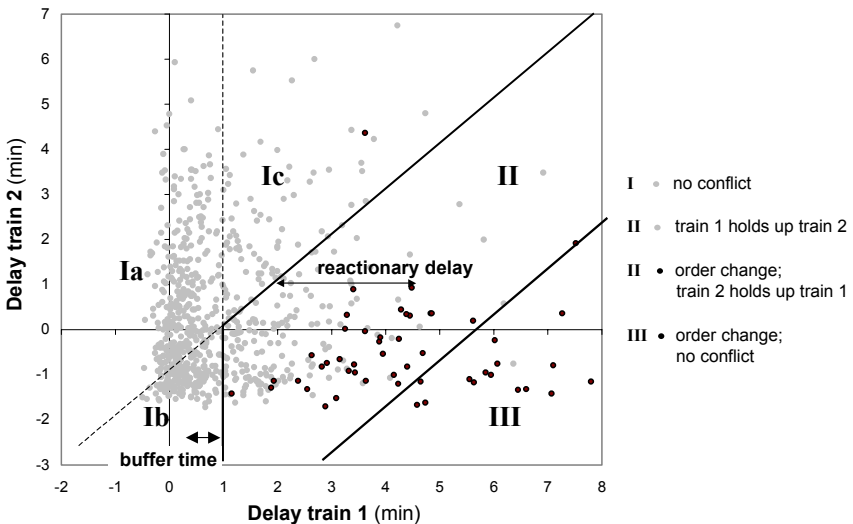


Figure 3: Determining reactionary delay from initial delays train 1 and train 2.



The observed percentage of reactionary delayed trains can well serve as an estimation of the probability of conflicts in the unchanged future situation. Figure 3: also gives an estimation of the time a train loses compared to an unperturbed run. The reactionary delay of a perturbed train 2 can be calculated as the horizontal distance between a category II dot and the left border line of category II:

$$r_2 = d_1 - \max(0, d_2) - b_{12} \quad (1)$$

with:

r_i reactionary delay to train i

d_i initial delay to train i

b_{ij} buffer time between train i and train j

In category II, both trains approach the junction at the same time. A conflict for either train is then inevitable. In case of train order change, the reactionary delay of train 1 can be calculated analogously to (1) as the horizontal distance to the right diagonal line. This delay is not considered here because its size is negligible for the entirety of “trains 1”. In case both trains come from the same track, order change is no option at all.

4.3 Relationship buffer time – reactionary delay

All calculations so far were observations, but the model is suitable to make predictions as well. Again, the conflict at Den Bosch serves as an example. Suppose that the departure of train 1 is put forward by one minute. This earlier departure enlarges the probability that train 1 has released the junction when train 2 is coming. Now, two assumptions are made:

- The delay distributions are independent. This only holds if the involved trains come from different tracks. If the trains have been following each other, train 1 may have been holding up train 2 before.
- The delay distributions stay constant. This is imaginable if the entire path of train 1 is advanced by one minute. Then, all “trains 1” depart one minute earlier from Den Bosch than they did in the timetable 2004.

Separating trains in time by one more minute corresponds with adding one minute of buffer time. Current buffer time can be seen left at the bottom in Figure 3:. By increasing this distance (Figure 4:: shifting the diagonal lines horizontally), the effect of a longer buffer time can be determined:

- Category I widens, i.e. a larger number of “trains 2” will not suffer any reactionary delay.
- The reactionary delay of perturbed trains (category II) will fall.
- The number of order changes will decrease; in some category II cases this will *create* a delay for train 2. In category II the model assumes that the train dispatcher always chooses the order with minimal total delay.



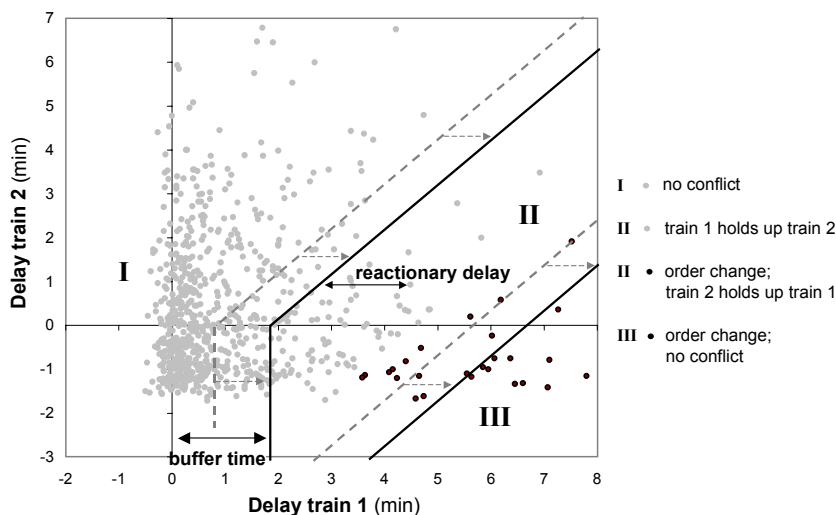


Figure 4: Determining reactionary delay at changing buffer time.

On the basis of these new figures, the average reactionary delay of all “trains 2” can be calculated. This was done for every value of the buffer time. This yields a relationship between buffer time and reactionary delay (see Figure 5; the black curve). Such a calculation was carried out for other short scheduled train successions in the case studies; each curve in the figure represents the relationship of a specific conflict. Although the reactionary delay depends highly on the delay distributions of both trains, the curves look similar: buffer time limits reactionary delay, but the benefits decrease with increasing buffer time. With a buffer time over 2 minutes the average reactionary delay becomes negligible.

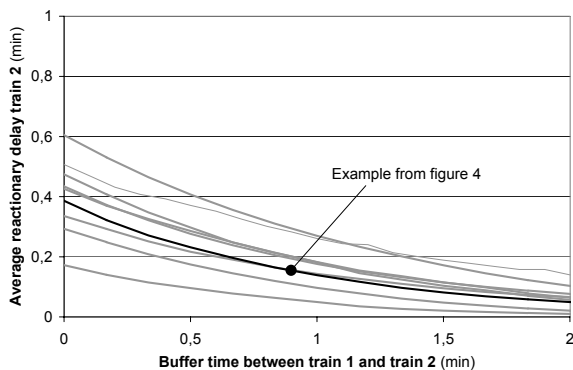


Figure 5: Calculated relationship between buffer time and reactionary delay.



5 Joint standard

Robust design standards should basically enable traffic to run smoothly under daily stochastic conditions. Following that principle this section will stepwise integrate running, dwell, and headway times toward a joint design standard.

5.1 Integration of running and dwell times

In sections 2 and 3 it was found that inter-city train's running times and dwell times at main stations are feasible. Local trains' paths, however, were shown to be too tight. This difference is caused by two elements:

- Variation in running times occurs around stops, of which local trains have more.
- The standard for line-side stops is too short.

These problems are tackled with two changes in the standards. First, equally distributed supplements (the current 7%) are replaced by a fixed supplement per main station stop (or: per node-to-node run) to cover acceleration and braking variation. The empirical data of inter-city trains indicates that 1.5 minutes per stop is sufficient for a running time feasible for most trains (over 75% of the unperturbed runs of all series in both case studies). To make a practicable standard, this value is also used for local trains per main station stop.

For local trains, secondly, line-side stops and running time variations around them are addressed in connection. The data shows that both components together require an extra 0.4 minute per stop, on top of the 1.5 minutes mentioned above. An easy way to include that time in the train path is adding it to the dwell time at line-side stops: it does not matter whether margins are put in running times or in dwell times, whenever the *sum* of both is large enough. With a current standard of 0.6 minute on average this makes up a new dwell time of 1.0 minute.

5.2 Integration of running and headway times

The previous paragraph gave an integrated approach for running and dwell times. With that, feasible paths can be planned for autonomously running trains. When multiple trains run on a network, they may interfere and cause conflicts. Here headway times show up. The idea is this:

- Conflicts are *prevented* with buffer times between train paths.
- The reactionary delay still occurring is *cured* with an extra running time supplement for the succeeding train. This supplement is equal to the expected average delay, according to Figure 5:.

When consistently network-wide implemented, this regime will generate a planning which is robust against daily variations. A usual requirement for robustness is a not-increasing average delay [3] and that is achieved with this standard. The quantified interchangeability of different types of margins provides flexibility in timetable design. An appropriate combination of running time and headway time should be chosen situation-specifically.



5.3 Overall integration

On the basis of the steps described above, a joint design standard for running, dwell and headway times is suggested (under other circumstances, values may differ but the concept still holds):

- Between two main nodes trains (both inter-city and local ones) have a fixed supplement of 1.5 minutes covering acceleration and braking variations around the nodes, on top of the minimal running time.
- The dwell time standard at line-side stops is raised from 0.5-0.7 minute to 1.0 minute in order to cover excessive dwelling and varying driving behaviour around these stops. Dwell times at the nodes stay as they are.
- At short headways, a succeeding train gets an additional running time supplement depending on the buffer time between the train movements, compensating for reactionary delay. This supplement varies from 0.4 minute in case of no buffer to zero in case of 2.0 minutes buffer time.

6 Conclusions

A railway timetable needs certain margins in order to deal with operational variation. Computer tools enable detailed analysis of train running data, which shows that design standards for local services are too tight and that current planning practice insufficiently takes headway conflicts into account.

An alternative joint design standard is proposed, no longer determining running time supplements as a percentage of minimal running times, but allocating them where they are needed: depending on the number of stops. In conjunction, dwell times at line-side stops are extended. Conflicts between trains can be either prevented by separating paths with buffer times or cured by an additional supplement in the succeeding train's running time, compensating the expected reactionary delay. The suggested standard would improve punctuality of the railway service and flexibility of the timetable design process.

Acknowledgement

This project has been performed within the TRANSUMO program "Reliability of transport chains", funded by Dutch government and in co-operation with ProRail.

References

- [1] Fénix, J., "User centred design applied to increase timetable stability". Accepted for the 36. Tagung Moderne Schienenfahrzeuge, Graz, Sept. 2005
- [2] Goverde, R.M.P. en I.A. Hansen, "TNV-Prepare: Analysis of Dutch railway operations based on train detection data". In: J. Allen *et al*, Computers in Railways VII, WIT Press, Southampton, 2000, pp. 779-788
- [3] Union International des Chemins de fer, Capman Working Group, Capacity Management (Capman Phase 3). Summary report, Dec. 2004



***RTCSIM*: an innovative, extendable computation engine for timetable validation**

T. Polzin & R. Gooßmann
HaCon GmbH, Hanover, Germany

Abstract

In the beginning of 2005, HaCon started an initiative to redesign the computation engine for running times and mapping of train service dynamics as provided within the *TPS Timetable Planning System*, which is the core production system for railway operational planning in Denmark. In the light of increasing demands on timetable validation technology, this activity aimed at providing a future-safe solution for the integration of such technology into operational planning applications.

Meanwhile, the first versions of the newly designed module *RTCSIM* have been integrated within the actual software version of *TPS*. The system is used for integrated production planning over different time horizons including STP (Short Term Planning) at *Banedanmark*, *DSB* and *DSB S-train* as well as *Trafikstyrelsen*, the Danish regulation and certification body. It was implemented within the years 1999–2002 and has evolved steadily until today, facing imminent changes in business processes and system interfaces.

TPS features a high resolution level of infrastructure data including tracks and routes as well as detailed security system information laid out in a principle track plan. Train services are mapped into the model using detailed occupation and release information for the infrastructure elements following the UIC 406 leaflet. It includes conflict detection and resolution algorithms being applied to the overall capacity plan maintained by the system, featuring train services, temporary speed restrictions, track blockages or other withdrawals of network capacity. Using the modular architecture of the new computation engine will unleash several innovative potentials. This especially facilitates the future use of extensive functionality for assessment and optimisation of railway network capacity within *TPS*. With that, HaCon now continues with a long tradition of application development such as the well known *UX-SIMU* software package as supplied until 1999.

This paper gives an overview about the underlying concept and mechanisms of the new computation engine *RTCSIM* as well as an assessment of impacts on future timetable validation techniques.

Keywords: *timetable planning, operational planning, timetable validation, microscopic infrastructure, simulation, capacity assessment.*



1 Introduction

Recent assessments about using stochastic, synchronous timetable simulation as validation method within the annual timetable development process made clear that availability of complete and effective plans is a key issue (see e.g., Watson [8] and Weits [9]). Consequently, these validation methods must feature a high measure of integration into operational planning systems being used for management of the accurate and up-to-date “master plan”. Otherwise, loss of acceptance and thus, loss of timetable quality would be the result, especially when it comes to short term planning (STP) where the amount of daily service variance is high.

This paper outlines the process of implementation of the *RTCSIM* computation engine for timetable validation purposes, its underlying concepts and impacts on future timetable validation techniques.

The paper is structured as follows: In Section 2, we refer to the general business background of timetable planning in order to depict the need for a fast and reliable validation technique as part of the daily operational planning. Section 3 describes the implementation process of *RTCSIM*, starting with an overview of the goals related to this development initiative. Section 4 includes a more detailed depiction of software concepts and architecture including integration options of *RTCSIM*. In Section 5, we discuss the current status of the project followed by an introduction to two major innovative concepts of the *RTCSIM* module. Section 6 provides our brief conclusions.

2 Background

While pursuing their aim to deliver timetables of good quality, planners have to face several business process challenges, such as:

- The need for fast decisions on TOC (Train Operating Company) requests;
- Existing timetable regularity requirements and commercial impacts (e.g., performance guarantees);
- Increasing network capacity usage especially in station areas;
- The need for short term changes;
- The planning and operational rules to be obeyed;
- The train service request compliance; and
- Assumptions and existing knowledge about reality;

Dealing with these issues means dealing with plans that make use of limited resources and at the same time, ensure that these plans are feasible.

More practically, such plans feature not only realistic timetables but also the ability to overcome potential obstacles in operation as fast as possible. This ability is also referred to as the *timetable quality*. Providing high quality timetables while fulfilling the business process requirements as mentioned above makes it obvious that there is an immanent need for a fast and reliable decision support for immediately assessing the quality of changed or amended timetables as part of the daily planning tasks.



The ***TPS Timetable Planning System*** is the core production system for railway operational planning in Denmark, being in production under the name “*STRAX*” since 2002 (see also [2] for details). It makes use of a high resolution geography information including railway infrastructure on signal berth level and different signalling systems with their intrinsic properties. Train services are mapped into the model using detailed occupation and release information for the infrastructure elements following the UIC 406 leaflet.

The system includes an integrated kernel module for running time calculation and animation of train services of the maintained timetables which is based on the well-known *UX-SIMU* tool as supplied by HaCon until 1999. This tool and its synchronous timetable simulation was used at several railway customers including the German railways through-out the 1990’s for evaluation of railway infrastructure, timetables and operation.

2.1 *RTCSIM*

RTCSIM is designed as a separate computational module for timetable validation. It provides the following services:

- **Runtime Calculation (RTC):** The computation of one train running separately, using detailed information about trains and tracks; includes station track adjustment, trackway search, and the calculation of signalling system dependant occupation and conflict information.
- **Synchronous Simulation (SIM):** In this context, simulation means the representation of interacting train runs on a given infrastructure at a given time interval. Typically it is run under perturbed conditions (e.g., stochastically introduced delays or defects), such that signalling systems’ safety rules, dispatching rules (e.g., prioritisation, unscheduled overtaking, re-routing), and functionality to avoid unrealistic situations (in particular, deadlocks) are applied. *RTCSIM* provides simulations in a special form of *Simulation Event Protocols (SEP)*, which will be described in Section 4.4 together with basic analytical facilities that can be used by the visualisation front end.
- **Additional Services:** E.g., for basic queries of driving dynamics, such as for the “maximum permissible load” of a train run.

3 Implementation process

3.1 Goals of the *RTCSIM* development initiative

The main goals for the development have been the following:

- **Compatibility with the existing *UX-SIMU* kernel module**
As the existing *UX-SIMU* based kernel module has a profound empirical basis and the *TPS* customers maintain a huge amount of data within their *TPS* systems, it has been the aim to provide compatibility to the existing kernel module concerning the calculation results as well as regarding the usability of available *TPS* data.



- *Extensibility*

In particular, for supporting new signalling systems (see Section 4.3) and evolving dispatching functionality (see Section 5.1).

- *Maintainability*

As *RTCSIM* will be used in a commercial product with long lifetime, it is a high priority goal to have modules that are well structured and comprehensible in the beginning and also maintain their structure after changes and additions.

- *Quality Assurance*

Being a part of the timetable production system at DSB and Banedanmark, the requirements for quality assurance (fault tolerance, quality of results, stability of product version releases) needed to be fulfilled.

3.2 Setting up the project

The *RTCSIM* project has been started in the beginning of 2005. The project schedule included milestones for delivery of different sub-modules in different stages as well as intensive testing and benchmarking activities. Emerging from a long history of providing planning systems and simulation tools for the railway industry, the in-house knowledge about relevant needs and methods formed the base for the set-up of requirements. Moreover, recent results especially from the fields of operational railway science and software engineering were incorporated into the specification documents used for the *RTCSIM* development. The development process was characterised by tight contact to railway planners, scientists and users of simulation models throughout the complete realisation period.

3.3 Involved methods and railway science

The process used an in depth analysis of the old *UX-SIMU* rooted computation in *TPS*. We evaluated the operational railway science background and re-thought the mathematical formulas for the driving dynamics. Additional input from conferences and journals was taken into account.

4 Software design concepts

In this section, we describe the main software design concepts that have been applied to achieve the goals as mentioned in Section 3.1.

4.1 Unification

The *UX-SIMU* based kernel module had multiple routines for similar things, e.g., the computation of one train on its own and of multiple trains at once used different concepts for representing the security system aspects. The new *RTCSIM* module has just one routine for one thing. For example, it adopts the paradigm “a runtime calculation is the simulation of one train running on its own”. This leads to significantly improved maintainability and quality assurance.



4.2 Quality assurance

For *RTCSIM* we use different automatic testing procedures that are performed and checked every night.

- Unit Tests: Test for the basic functional units, tested on their own, typically on restricted test data, immediately detecting if a new development breaks already implemented functionality. This encourages developers to refactor parts of the program (see [1]).
- Integration Tests: Tests for the integrated functionality, comparing just the over all results.
- Regression Tests: Computer programs tend to reproduce bugs that have been fixed. Hence, test cases that showed the previously buggy behaviour have some indication to represent critical cases. Therefore, we implemented a system that checks these possibly critical cases.
- Performance Tests: Automatic performance tests give an immediate feedback on the performance implication of program changes.

Additionally, we use a report on code coverage showing which parts of the code are not covered by any test as well as a memory analysis report based on *valgrind* [6].

4.3 Modularization

In the programming of the existing UX-SIMU based computation engine the aspects dispatching, driving dynamics, and security were interwoven. In *RTCSIM* these aspects are handled by separate modules.

This improves

- extensibility (because it is easier to replace one aspect of functionality by something else);
- maintainability (because the code has a clearer structure); and
- overall system quality (because the separate modules can be tested on their own more rigorously [3]).

4.3.1 Example: driving dynamics encapsulation

As an example, we outline the successful encapsulation of driving dynamics aspects. Here, the major invention is the development of the module interfaces. For flow of information into the driving dynamics module we use standardised *speed restrictions*, each of them containing information on when the speed restriction is transmitted to the train, and on which region it applies. For the flow of information back from the driving dynamics we use *driving triggers*. These triggers can be thought of virtual balises that perform some action as the train passes them (e.g., blocking some part of the infrastructure).

4.4 Introduction of Simulation Event Protocols (*SEP*)

Another major modularization step is the invention of Simulation Event Protocols (*SEP*). These protocols cover all information that is necessary to



represent a given point in time as part of the simulation time frame. Improvements through *SEPs* are:

- Quality Assurance: *SEP* is a storable, hence manageable document with long term reproducibility, independent of forthcoming *RTCSIM* versions. It includes quality assurance information such as, e.g., creator, creation timestamp, source project etc.;

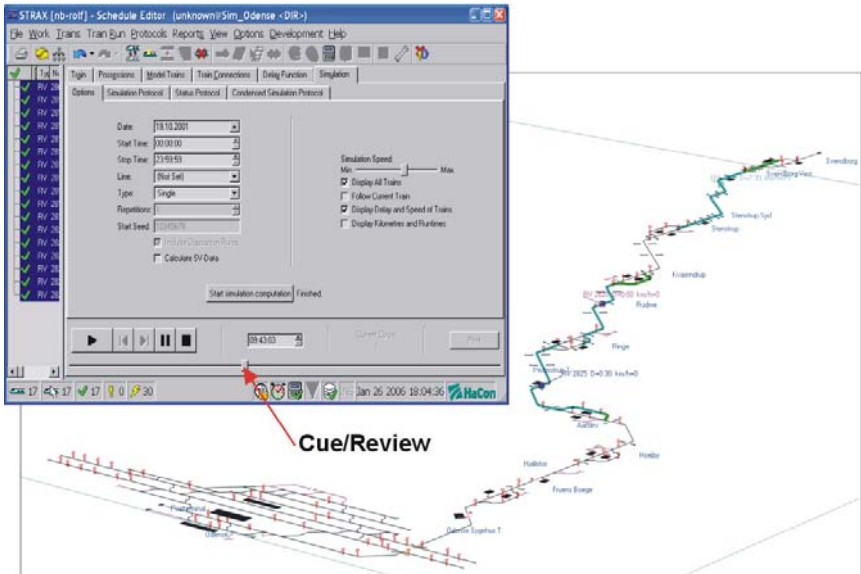


Figure 1: RTCSIM Simulation with reviewing feature.

- Improved Analysis: Critical situations can be examined over and over, using a tape recorder like cue/review functionality (going backwards in time); see also figure 1. A critical situation may be recorded as movie to show it even without a *TPS* application in the background;
- Keyfigure Tracking: Not only train animation can be seen repetitively but also the changes of absolute or accumulated keyfigures such as number of hindrances, sum of delays, capacity consumption etc.;
- Communication: *SEPs* can be sent easily via e-mail;
- More effective utilisation of computational power:
 - No need to re-compute existing computations;
 - Task scheduling, i.e., distribution of simulation jobs on multiple processors (scalability);
- Software Design Aspects:
 - Thoroughgoing separation of GUI-visualisation and simulation;
 - *SEP* files are available in binary, structured text or XML format;

4.5 Integration options

As the *RTCSIM* software is designed to be a separate module with a well defined interface on binary/ XML basis, there are several integration possibilities.

- *RTCSIM* can be used inside the *TPS* GUI or as command line interpreter in foreground or background mode. Each integration makes use of a *TPS-RTCSIM*-interface module that connects *RTCSIM* with *TPS* data.
- Just replacing this interface module, *RTCSIM* could be connected to other applications' data or work inside another application, if the data model of the other application maps somehow to *RTCSIM*'s needs for computation.

5 Current state and perspectives

Meanwhile, the main goals as mentioned in Section 3.1 have been reached and the first versions of the newly designed module *RTCSIM* has been integrated within the actual software version of *TPS*.

Concerning the computation speed, *RTCSIM* outperforms its predecessor module by a factor of 2 to 3. The clean architecture of *RTCSIM* offers many potentials for optimisation. Realising these potentials is a currently ongoing process, therefore it is not sensible to give quantified performance tables here. In the following, we introduce two major innovative concepts of *RTCSIM*.

5.1 The concept of control laws

As the purpose of stochastic simulation is to deal with perturbed operations condition, the near realistic handling of these perturbations (i.e., by dispatching) is an important goal for simulation routines. The dispatching must be powerful, so that it can handle the situation appropriately, but also not better than reality, because otherwise scenarios that have even been simulated successfully could be impossible to handle in real life situation. From practical experience and reports in the literature [7] with simulation algorithms we derive another major requirement: *The dispatching decisions as being applied by the program must be presented to the user in a reasonable and understandable way.*

The basic approach is that there is a larger number of concrete Control Laws that are applied for concrete dispatching decisions, but they are not generated by hand, but by *Meta Control Laws* that specify the desired dispatching policies (like prioritisation) on a higher level. Making use of technically challenging algorithms, *Meta Control Laws* produce tables of Control Laws that can be reviewed, checked, corrected, or incremented manually. This significantly improves the comprehensibility of results compared to a "black box" dispatching algorithm.

5.1.1 Control laws

A Control Law in our context is a decision rule which describes an operational constraint or behaviour for prevention or treatment of operational conflicts.



It has the following basic structure:

If (OpCondition[Filter]) Then (OpAction[Filter]), where “OpCondition” describes an operational scenario involving one or more train services. “OpAction” is either “Insert (e.g., register a facultative train service)”, “Delete (cancellation of a train service)” or “Change (e.g., waiting time)” plan elements. OpCondition and OpAction may both be related to filters, restricting their validity in

- geography (e.g., region, line, station, operational point);
- time (e.g., operating days, daytime interval); and
- operation (e.g., train category, train operating company, delay, passenger occupancy degree).

If no time is specified for the OpAction filter, the OpAction is valid as long as the OpCondition is matched. This is especially useful when formulating rules for deadlock detection and prevention as suggested, e.g., in [4] and [5]. OpCondition and OpAction may also include threshold values and parameters as well as logic terms “and”, “or” and “not”. Another option is the use of projected track occupation information in order to derive forthcoming (knock-on) delays and conflicts, i.e., conditions such as “If (... and +delay at next stop > m minutes...)”.

Control Law example:

If (a regional passenger train A is delayed more than n minutes between station X and Y and a IC train B running behind it is hindered by A in off-peak hours)
Then (A shall have an extra stop in track t at station Y in order to let B pass).

In a similar way, all other typical train prioritisation rules can be formulated.

5.1.2 Meta Control Laws (MCL)

Since OpConditions and OpActions may be formulated using different levels of abstraction, it is useful to distinguish between Control Laws that can be directly applied to existing data entities and more general rules, which need entity instantiation and extension and by that, producing itself a set of concrete Control Laws. Such general rules we call *Meta Control Laws (MCL)*.

Meta Control Law example:

If (delayed regional passenger traffic causes hindrances to long distance traffic in off-peak hours) Then (overtaking should take place).

It is intended to make the use of Meta Control Laws as flexible as possible since this minimises the amount of manual work and the high abstraction level of *MCLs* ensures long term consistency disregarding detailed changes of plan, infrastructure or other planning data over different timetable periods.

Since by definition, the scope of *MCL* validity is not limited it is evident to use this new concept for official access and regulation rules as well as for complete network strategies. Moreover, such rules and strategies are easy to change in order to derive performance keyfigures for different scenarios.



5.1.3 Open API for dispatching plug-in

Another approach for powerful, predictable, comprehensible dispatching decisions is the supply of a C++ program interface (API) for dispatching: Our concept includes the provision of information services (based on the current simulation situation and the planned schedule) to an external software plug-in that decides on dispatching questions like unscheduled overtaking. Alternatively, a user-editable script language (LUA, Python) can be used for this purpose, enabling rapid configuration of dispatching tasks.

5.2 The restarting mechanism

As the simulation event protocol (*SEP*) contains all information to display a simulation at a certain point in time, it also contains the information to *restart* a simulation at a certain point in time, possibly with modified parameters. This enables us to look at possible applications:

5.2.1 Interactive refinement of dispatching

It is easier to tune the dispatching algorithm, if one can look at a certain situation, where something went wrong. The user could manipulate the Control Laws and restart the simulation at this situation. This cycle can be repeated quickly until the desired behaviour is achieved.

5.2.2 Automatic refinement of dispatching

In case of deadlocks or other undesired dispatching effects, the simulation could use backtracking to find a feasible solution.

5.2.3 Real time dispatching

If there is a possibility to receive (partial) information about the current train positions, these information can be merged into the simulation at this point in time. Together with the possibility to adjust dispatching decisions interactively (see Section 5.2.1) this would lead to a powerful support tool for real time dispatching.

6 Conclusions

We described the background, the main goals and some of the underlying concepts for successfully building *RTCSIM* as an extendable computational engine for timetable validation. We believe that the implementation of these concepts was very successful, leading to a future-safe basis for further development, for which we gave a short overview.

One of the most important innovations of *RTCSIM* lies in the ability of applying time reversion in simulation visualisation and re-entry into retrospective simulation events (backtracking). This ability can be used for pre-emptive dispatching rules taking prospective delay or conflict situations into account.



Our approach includes the provision of a Control Law plug-in for *RTCSIM*, which represents an open interface for external experts for defining and testing dispatching procedures on their own.

We introduced Meta Control Laws (*MCL*) as a major concept for mapping of high level rules for regulation and network strategy. *MCLs* are supposed to serve as a major planning instrument of forthcoming simulation applications.

The XML based *RTCSIM* communication interface facilitates easy integration in other operational planning system environments.

References

- [1] Beck, K., "Test Driven Development", *Addison-Wesley*, 2002, Boston.
- [2] Kaas, A. H. & Gooßmann, R., "Implementation of the Timetable Planning System STRAX/TPS in Denmark", *COMPRAIL 2004*, Dresden.
- [3] Lakos, J., "Large Scale C++ Software Design", *Addison-Wesley*, July 1996, Boston.
- [4] Pachl, J., "Safe disposition and scheduling in railway operation", *Signal + Draht* (92) 5/200, p. 38-41.
- [5] Pachl, J., "Steuerungslogik für Zuglenkanlagen zum Einsatz unter stochastischen Betriebsbedingungen", *Institut für Verkehr, Eisenbahnwesen und Verkehrssicherung der TU Braunschweig*, Heft 49, 1994.
- [6] *Valgrind*: www.valgrind.org.
- [7] C.R. Warninghoff, C. Ferchland, "Nutzung von Simulationen zur Unterstützung der betrieblichen Infrastrukturplanung", *ETR* no. 53 (2004), vol 7/8, p. 490-498.
- [8] Watson, R., "Using stochastic simulation to predict timetable performance – status and developments in the UK", *IAROR 2005*, Delft.
- [9] Weits, E.A.G., Zinsmeister, E.R., "On the validity of simulation models of railway operations, a plea for standardised models", *IAROR 2005*, Delft.



Evaluating stochastic train process time distribution models on the basis of empirical detection data

J. Yuan, R. M. P. Goverde & I. A. Hansen
*Faculty of Civil Engineering and Geosciences,
 Delft University of Technology, The Netherlands*

Abstract

This paper evaluates several commonly applied probability distribution models for stochastic train process times based on empirical data recorded in a Dutch railway station, The Hague Holland Spoor. An initial guess of model parameters is obtained by the Maximum Likelihood Estimator (MLE). An iterative procedure is then followed, in which large delays are omitted one by one and the distribution parameters are estimated correspondingly using the MLE method. The parameter estimation is improved by minimizing the Kolmogorov-Smirnov (K-S) statistic where of course the empirical distribution is still based on the complete data set. A local search is finally performed in the neighbourhood of the improved model parameters to further optimize the estimation. To evaluate the distribution models, we compare the K-S statistic among the fitted distributions with optimized parameters using the one-sample K-S goodness-of-fit test at a commonly adopted significance level of $\alpha = 0.05$. It has been found that the log-normal distribution can be generally considered as the best approximate model among the candidate distributions for both the arrival times of trains at the platform and at the approach signal of the station. The Weibull distribution can generally be considered as the best approximate distribution model for non-negative arrival delays, departure delays and the free dwell times of late arriving trains. The shape parameter of the fitted distribution is generally smaller than 1.0 in the first two cases, whereas it is always larger than 1.0 in the last case. These distribution evaluation results for train process times can be used for accurately predicting the propagation of train delays and supporting timetable design and rescheduling particularly in case of lack of empirical data.

Keywords: train delays, running and dwell times, track occupancy times, statistical distribution, the K-S test.



1 Introduction

Modelling the distribution of train process times is an important research topic. Arrival and departure delay distributions reflect the punctuality level of trains at stations. Based on the distribution of input delays of trains at the boundary of a railway network and the distribution of primary delays within this network, the distribution of knock-on delays and that of the resulting exit delays can be estimated, which supports timetable design and operations management [10].

Train process time distributions are often assumed based on experiences from real operations and limited literature [2], [3], [6], [8], [9] exists with respect to statistical inference of the distributions using empirical data observations. Track occupation and release records show the total delays of trains and may include knock-on delays. Therefore, data filtering is necessary to fit the distribution of primary delays on the basis of train detection data [9]. To the best of our knowledge, there is no publication that evaluates the conditional train running and dwell and track occupancy time distributions in the case of different aspects of relevant block signals, which can be used for estimating knock-on delays more accurately [10].

This paper evaluates several commonly applied distribution models for stochastic train process times on the basis of empirical traffic data recorded in a Dutch railway station The Hague Holland Spoor (The Hague HS). The evaluation of distribution models is performed not only for the arrival times (delays), non-negative arrival delays and departure delays of trains at the station, but also for the arrival times of trains at the boundary of the local railway network and the train running, dwell and track junction occupancy times within the local network. This paper is structured as follows. Section 2 outlines the distribution models to be evaluated and the evaluation method. The results of the distribution evaluation for the process times of trains are then discussed in Section 3. Finally, the main conclusions are drawn in Section 4.

2 Distribution models and the evaluation method

A variety of theoretical probability distributions, including the normal, uniform, exponential, gamma, beta, Weibull, and log-normal distributions [5], have been adopted in the literature to model the stochasticity of train process times. Given a continuous distribution model, the parameters are estimated on the basis of empirical data using e.g. the maximum likelihood method. The resulting fit can be tested using the Kolmogorov-Smirnov (K-S) goodness-of-fit test. This test is based on the K-S statistic, defined as the maximum absolute difference between the empirical and fitted cumulative distribution function [5]. However, the estimated parameters are generally sensitive to outliers in the data set. We therefore used an iterative parameter estimation method. An initial guess of the model parameters is obtained using the Maximum Likelihood Estimator (MLE) of the complete data set. Next, the large delays in the original data set are omitted iteratively one by one estimating the distribution parameters correspondingly using the MLE method. In each iteration, we compute the K-S



statistic where of course the empirical distribution is still based on the complete data set. The iterative procedure terminates if the K-S statistic cannot be decreased any more. After the iterative procedure, we apply a local search in the neighbourhood of the parameter estimate to further optimize the parameter estimation by minimizing the K-S statistic. It should be mentioned that the location parameter of each distribution model except for the normal distribution was taken as the minimum value of the empirical data observations.

To evaluate the candidate distributions, we compare the K-S statistic among the fitted distributions with optimized parameters using the one-sample K-S test [5] at a commonly adopted significance level $\alpha = 0.05$. To visualize the quality of distribution fitting for the process times of trains, we compare the fitted distribution density curve with the kernel estimate and empirical histogram [7] and apply the distribution differences plot [1] for the fitted distribution and the empirical one.

3 Evaluation results

The distribution of train process times may depend on the types and routes of trains. We hence evaluate the distribution models for the process times of trains per train series in both the southbound and northbound directions at The Hague HS railway station.

3.1 Arrival times

The modelling of train arrival time distribution is a prerequisite for predicting the propagation of train delays at stations. To incorporate the impact of the knock-on delays caused by route conflicts in a delay propagation model, we need to distinguish the arrival times of trains at the station platform from that at the approach signal of the station. Early arriving trains are often considered as punctual trains in some delay propagation models [4], where the distribution of non-negative arrival delays is of the main concern.

It has been found that the location-shifted log-normal distribution is the best approximate model among the candidate distributions in 9 and 11 of the 14 considered cases for both the arrival times of trains at the approach signal of the station and at the platform track, respectively. This distribution model has not been rejected by the K-S test in 9 and 10 of the 14 cases for both the arrival times. The optimized parameters and the K-S test results for the log-normal distribution are given in Table 1, where μ and σ represent the mean and standard deviation of the underlying normal distribution and p denotes the p-value [5] of the K-S test.

Figure 1: shows the optimized log-normal distribution density curve, kernel estimate and empirical histogram for the arrival times of the northbound intercity train series IC2100N at the approach signal of the station. The corresponding distribution differences plot is shown in Figure 2; where the two dotted horizontal lines represent the critical error bounds for the K-S test. In both figures, the reference time is defined at the scheduled arrival time of the studied



train series at the station. The optimized log-normal fit matches well the empirical data. In this case, the distribution differences plot does not cross the error bounds, therefore the location-shifted log-normal distribution has not been rejected by the K-S test.

Table 1: Optimized parameters and the K-S test results of the log-normal fit for the arrival times of trains in The Hague HS.

Train Series	At the approach signal				At the platform track			
	μ	σ	Best	p	μ	σ	Best	p
IR2200S	4.6	0.6	×	0.03	4.8	0.6	√	0.16
IC2100S	4.3	0.7	√	0.61	4.5	0.7	√	0.38
IC2400S	4.2	0.8	√	0.25	4.4	0.9	√	0.42
INT600S	4.9	0.7	×	0.18	5.0	0.5	√	0.20
HST9300S	4.8	0.9	×	0.01	4.8	0.9	×	0.00
AR5000N	4.7	0.4	√	0.00	4.7	0.5	√	0.00
AR5100N	4.5	0.5	√	0.00	4.6	0.5	√	0.00
IR2200N	4.6	0.4	×	0.00	4.5	0.5	×	0.00
IC1900N	4.9	0.6	×	0.15	5.1	0.5	×	0.16
IC2100N	5.0	0.7	√	0.55	5.0	0.7	√	0.14
IC2400N	4.8	0.6	√	0.31	4.9	0.6	√	0.79
IC2500N	4.8	0.6	√	0.13	4.9	0.5	√	0.15
INT600N	4.8	0.7	√	0.26	4.9	0.6	√	0.22
HST9300N	4.6	0.7	√	0.38	4.6	0.8	√	0.30

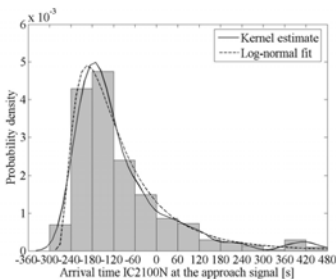


Figure 1: Log-normal fit, kernel estimate and histogram for the arrival times of IC2100N at the station approach signal.

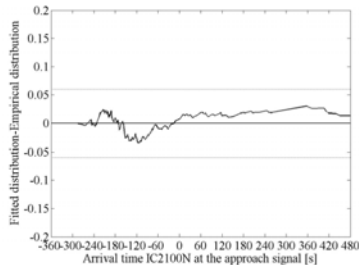


Figure 2: Distribution differences plot for the log-normal fit and the arrival times of IC2100N at the station approach signal.

We have also evaluated the candidate distributions for non-negative arrival delays of trains at the platform track. The Weibull distribution is the best fit among the candidate distributions in 7 of the total 18 studied cases. This distribution model has not been rejected by the K-S test in 17 of the 18 cases. Since the exponential distribution has been widely used to model the

stochasticity of non-negative arrival delays [2], [6], [9], we visualize the goodness-of-fit of both the Weibull and exponential distributions in Figure 3: and Figure 4: for the northbound intercity train series IC2100N. The non-negative arrival delays fit to the Weibull distribution with a shape parameter of 0.8 overall better than the exponential distribution especially at the range of larger delays. Both the distributions have not been rejected by the K-S test in this case.

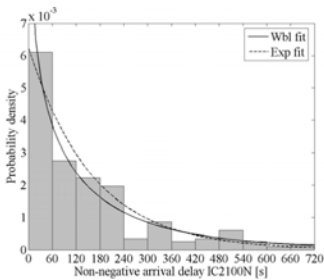


Figure 3: Fitted density curves and histogram of non-negative arrival delays of IC2100N.

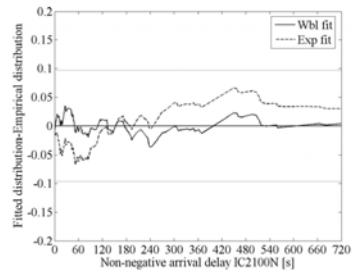


Figure 4: Distribution differences plots for non-negative arrival delays of IC2100N.

In conclusion, the log-normal distribution can be generally considered as the best model among the candidate distributions for both the arrival times of trains at the platform and at the approach signal of the station. The Weibull distribution matches well non-negative arrival delays. For simplicity, we may use the exponential distribution, which is a special type of the Weibull distribution, to be as an approximate distribution model for non-negative arrival delays if the density is decreasing.

3.2 Departure delays

Departure delays are non-negative since trains are not allowed to depart from the station earlier than the scheduled departure time. The distribution of departure delays can be used to predict the distribution of outbound track release times and the distribution of train arrival times at the following stations.

It has been found that the Weibull distribution is the best approximate model among the candidate distributions for the departure delays of trains in 11 of the total 18 studied cases. The exponential distribution is the best approximate model in 2 of the 18 cases. In addition, both the distributions have not been rejected by the K-S test in 10 of the total 18 cases. Figure 5: and Figure 6: visualize the goodness-of-fit of both the Weibull and exponential distributions for the departure delays of the southbound intercity train series IC2400S. Early arriving trains usually do not depart very late and some trains may arrive at and depart from the station very late, which results in a very steep histogram of the departure delays. The Weibull fit with a shape parameter of 0.8 matches the



delays overall better than the exponential fit. In this case, the former distribution has not been rejected while the latter distribution has been rejected by the K-S test. Thus, we can generally consider the Weibull distribution to be as the best approximate model among the candidate distributions for departure delays. Just like for non-negative arrival delays, the exponential distribution can be considered as an approximate distribution model for departure delays if the density is decreasing.

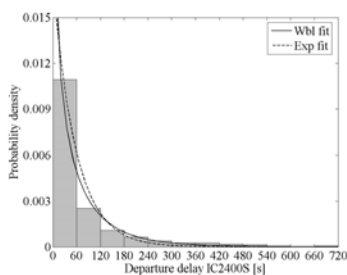


Figure 5: Fitted density curves and histogram of departure delays of IC2400S.

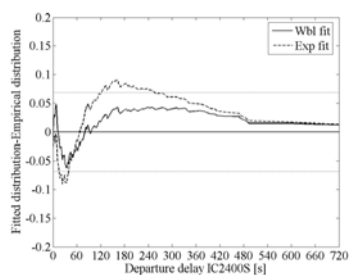


Figure 6: Distribution differences plots for departure delays of IC2400S.

3.3 Dwell times

The dwell times of trains are the difference between the arrival and departure times. Early arriving trains generally have much longer dwell times and they are not of our main concern. To estimate the knock-on delays and departure delays of trains at stations, it is critical to obtain the distribution of the free dwell times for late arriving trains [10]. The free dwell time of a train is defined as the necessary dwell time for passenger alighting and boarding in the absence of hindrance from other trains.

We fit the free dwell time distribution for late arriving trains per train series in one direction using the dwell time observations of the trains that satisfy

$$A_i > a_i \text{ and } C_i \leq A_i + 30$$

where, A_i and a_i denote the actual and scheduled arrival time of train i at the platform track and C_i represents the clearance time of the outbound route of this train. The unit of these times is in seconds. Late arriving trains are selected by the former inequality and the latter inequality ensures that the chosen trains are not hindered by other trains at the station after a minimal dwell time of 30 s.

It has been found that the Weibull distribution is the best approximate model among the candidate distributions for the free dwell times of late arriving trains in 16 of the total 18 studied cases. In addition, this distribution model has not been rejected by the K-S test in all the cases. Figure 7: and Figure 8: visualize



the goodness-of-fit of the Weibull distribution model with a shape parameter of 1.9 in case of the northbound interregional train series IR2200N. The fitted distribution matches well the kernel estimate for the empirical data and it has not been rejected by the K-S test. In conclusion, the Weibull distribution with a shape parameter larger than 1.0 is the best approximate model among the candidate distributions for the free dwell times of late arriving trains.

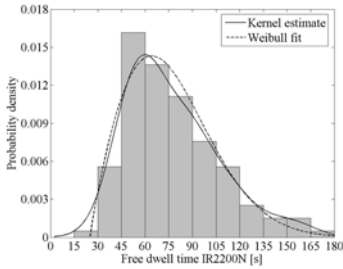


Figure 7: Weibull fit, kernel estimate and histogram for the free dwell times of late arriving trains of IR2200N.

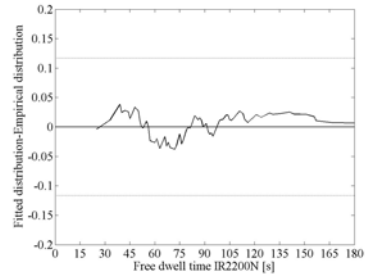


Figure 8: Distribution differences plot for the Weibull fit and the free dwell times of late arriving trains of IR2200N.

3.4 Running times and track occupancy times

The distribution of train running times and that of track occupancy times are required to estimate the propagation of train delays in a railway network. In this paper, we focus on statistical distribution of the running times of trains on the preceding block of The Hague HS station and that of the occupancy times of adjacent junctions around this station.

In case of an approaching train, if the inbound route is released earlier than the time of the train arriving at sight distance of the approach signal, the train approaches the station at the free running speed. Otherwise, the train is hindered and has to decelerate and even stop on the preceding block of the station. To accurately estimate the knock-on delays caused by route conflicts in a station area, it is necessary to investigate the conditional distributions of inbound train running and track occupancy times in the case of different aspects of the approach signal and home signal of the station. For a departing train, if it is hindered due to outbound route conflicts, it dwells at the station for a longer time, but running on the next track sections will not be hindered again. In this case, the conditional distributions are not applicable.

To model the conditional distributions of inbound train running and track occupancy times based on a statistical analysis of the empirical data, the first step is to classify the data observations. By comparing the arrival time of each train at the approach signal to the clearance time of the inbound route, we have extracted a data set suited for fitting the free train running and track occupancy



time distributions in each studied case. A hindered approaching train may pass the home signal with reduced speed without a stop or with acceleration speed after a stop in front of this signal. Since the standstill of a train on track is not recorded, we cannot directly identify whether or not a hindered train stops before the home signal based on track occupancy and release records. Adopting the k-means routine within the statistical analysis tool S-Plus [7], we have split the data sample of hindered trains for each studied train series into two separate parts which correspond approximately to the two cases mentioned in the above. However, for the hindered trains that stop before the home signal, it is still unknown when these trains stop. Therefore, we have lack of the running times of these trains on the preceding block of the station.

For the free running times of trains on the preceding block of the station, most of the candidate distributions have been rejected by the K-S test. Both the Weibull and normal distributions have not been rejected by the K-S test in 2 of the total 13 studied cases. In addition, each of these distributions is the best approximate model among the candidate distributions in 5 of the 13 cases. For inbound junction occupancy times by the free passing trains, the Weibull distribution is the best approximate distribution model in 3 of the total 4 considered cases and has not been rejected by the K-S test. The normal and Weibull distribution is the best fit among the candidate distributions for outbound track occupancy times in 2 and 1 of the total 3 considered cases, respectively. In addition, both the distributions have not been rejected by the K-S test in 2 of the 3 cases. The goodness-of-fit of the Weibull and normal distributions for the above-mentioned train process times is shown in Figure 9: and Figure 10:, respectively.

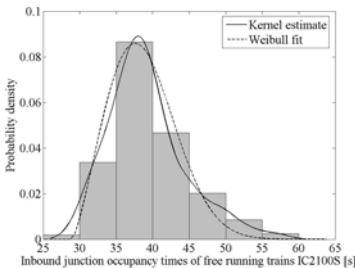


Figure 9: Weibull fit, kernel estimate and histogram for inbound junction occupancy times of free running trains IC2100S.

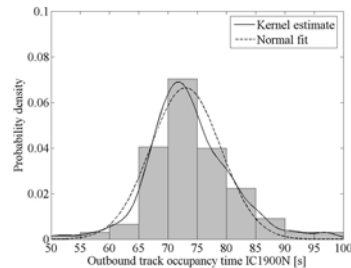


Figure 10: Normal fit, kernel estimate and histogram for outbound junction occupancy times of IC1900N.

In the case of the hindered trains that do not stop before the home signal, the Weibull and normal distribution is the best approximate distribution model for the running times on the preceding block of the station in 2 and 1 of the total 6 considered cases, respectively. In addition, both of the distributions have not been rejected by the K-S test in the 6 cases. For inbound junction occupancy



times by the hindered trains, the normal and Weibull distribution is the best approximate distribution model in 2 and 1 of the 5 considered cases and these distributions have not been rejected by the K-S test in the 5 cases. In the case of the hindered trains that stop before the home signal, the inbound junction occupancy times fit best to the normal distribution in the two considered cases and this distribution has not been rejected by the K-S test in one of the two cases. The goodness-of-fit of the Weibull and normal distributions for the above-mentioned train process times is given in Figure 11: and Figure 12:, respectively.

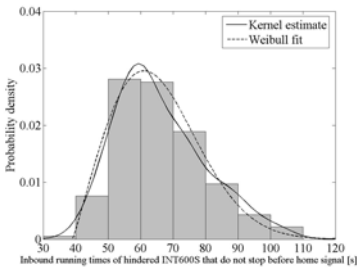


Figure 11: Weibull fit, kernel estimate and histogram for inbound running times of hindered INT600S that do not stop before the home signal.

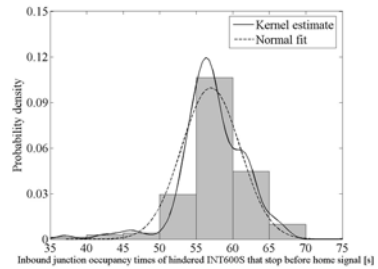


Figure 12: Normal fit, kernel estimate and histogram for inbound junction occupancy times of hindered INT600S that stop before the home signal.

In conclusion, it is difficult to find a good distribution for the conditional train running and track junction occupancy times in the case of different aspects of relevant block signals. This might be because of the big variation of train speed on the short track sections in the complicated station and junction area.

4 Conclusions

We have compared several commonly applied distribution models for train process times on the basis of empirical train detection data recorded at a Dutch railway station The Hague HS. It has been found that the log-normal distribution can be generally considered as the best approximate model among the candidate distributions for both the arrival times of trains at the platform and at the approach signal of the station. The Weibull distribution can generally be considered as the best approximate distribution model for non-negative arrival delays, departure delays and the free dwell times of late arriving trains. The shape parameter of the fitted distribution is generally smaller than 1.0 in the first two cases, whereas the shape parameter is always larger than 1.0 in the last case. For simplicity, the exponential distribution can be used as an approximate distribution model for non-negative arrival delays and departure delays if the density is decreasing.



Acknowledgement

This publication is a result of the research programme Towards Reliable Mobility, carried out within the Transport Research Centre of Delft University of Technology.

References

- [1] Averill, M.L. & Kelton, W.D., *Simulation Modeling and Analysis*, McGraw-Hill, 2000.
- [2] Goverde, R.M.P., Hooghiemstra, G. & Lopuhaä, H.P., *Statistical Analysis of Train Traffic: The Eindhoven Case*, DUP Science, Delft, 2001.
- [3] Hermann, U., *Untersuchung zur Verspätungsentwicklung von Fernreisezügen auf der Datengrundlage der Rechnerunterstützten Zugüberwachung Frankfurt am Main*, PhD thesis, Technische Hochschule Darmstadt, 1996.
- [4] Radtke, A. & Hauptmann, D., Automated planning of timetables in large railway networks using a microscopic data basis and railway simulation techniques, In: Allan, J. et al. (eds.), *Computers in Railways IX*, pp. 615-625, WIT Press, Southampton, 2004.
- [5] Ross, S. M., *Introduction to Probability and Statistics for Engineers and Scientists*, Elsevier, 2004.
- [6] Schwanhäusser, W., *Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn*, PhD thesis, RWTH Aachen, 1974.
- [7] S-Plus, *S-Plus 2000 Guide to Statistics*, Vol. 2, Data Analysis Products Division, Mathsoft, Seattle, 1999.
- [8] Steckel, J., *Strategische Optionen für die Zufällige Fahrzeit im Eisenbahnbetrieb*, PhD thesis, Hochschule für Verkehrswesen 'Friedrich List' Dresden, 1991.
- [9] Wendler, E. & Naeurig, M., Statistische auswertung von verspätungsdaten, *Eisenbahningenieurkalender EIK*, pp. 321-331, 2004.
- [10] Yuan, J. & Hansen, I.A., Optimizing Capacity Utilization of Stations by Forecasting Knock-On Train Delays, In: Hansen, I.A. et al. (eds.), *Proceedings of 1st International Seminar on Railway Operations Modelling and Analysis*, Delft, 8-10 June, 2005.



Section 11

Operations quality

This page intentionally left blank

Practical use of the UIC 406 capacity leaflet by including timetable tools in the investigations

A. Landex¹, A. H. Kaas², B. Schittenhelm³ & J. Schneider-Tilli⁴

¹*Centre for Traffic and Transport,
Technical University of Denmark, Denmark*

²*Atkins Denmark A/S, Denmark*

³*Rail Net Denmark, Denmark*

⁴*The National Rail Authority, Denmark*

Abstract

This paper describes the relatively new UIC 406 method for calculating capacity consumption on railway lines. The UIC 406 method is an easy and effective way of calculating the capacity consumption, but it is possible to expound the UIC 406 method in different ways which can lead to different capacity consumptions. This paper describes how the UIC 406 method is expounded in Denmark.

The paper describes the importance of choosing the right length of the line sections examined and how line sections with multiple track sections are examined. Furthermore, the possibility of using idle capacity to run more trains is examined.

At the end of the paper a method to examine the expected capacity utilization of future timetables is presented. The method is based on the plan of operation instead of the exact (known) timetable.

Keywords: capacity, capacity analysis, compressed timetable graphs, timetable, UIC 406, railway, railway operation.

1 Introduction

The UIC 406 leaflet from year 2004 [9] describes a simple, but fast and effective way to evaluate the capacity utilization of railway lines. The capacity analyses carried out during the last years using the UIC 406 method have been presented in a number of papers (e.g. [3] and [10]). However, it is possible to expound the



UIC 406 method in different ways which can lead to different results. In spite of that fact, hardly any analyses of the differences have been carried out.

This paper describes the UIC 406 method (cf. section 3) and a number of analyses carried out to generate a Danish consensus of expounding the UIC 406 leaflet for capacity analysis. A number of different analyses on relevant Danish railway lines have been carried out including:

- The length of the railway section examined (cf. section 4)
- The allowance of changing tracks on railway lines with multiple track sections (cf. section 5)
- The possibility of using idle capacity to run more trains (cf. section 6)
- The possibility of using the plan of operation instead of the exact timetable (cf. section 7)

At the end of the paper some conclusions and perspectives of the UIC 406 method are listed (in section 8).

2 Definitions

This paper uses terminology usually used in the railway literature. However, since the terminology differs from country to country, an overview of the terminology used in this paper is provided in Table 1.

Table 1: Short description of terminology.

Term	Explanation
Block occupation time	The time a block section (the length of track between two block signals, cab signals or both) is occupied by a train
Buffer time	The time difference between actual headway and minimum allowable headway
Headway distance	The distance between the front ends of two consecutive trains moving along the same track in the same direction. The minimum headway distance is the shortest possible distance at a certain travel speed allowed by the signalling and/or safety system
Headway time	The time interval between two trains or the (time) spacing of trains or the time interval between the passing of the front ends of two consecutive (vehicles or) trains moving along the same (lane or) track in the same direction
Running time supplement	The difference between the planned running time and the minimum running time

Some of the terms described in Table 1 are further illustrated in Figure 1.



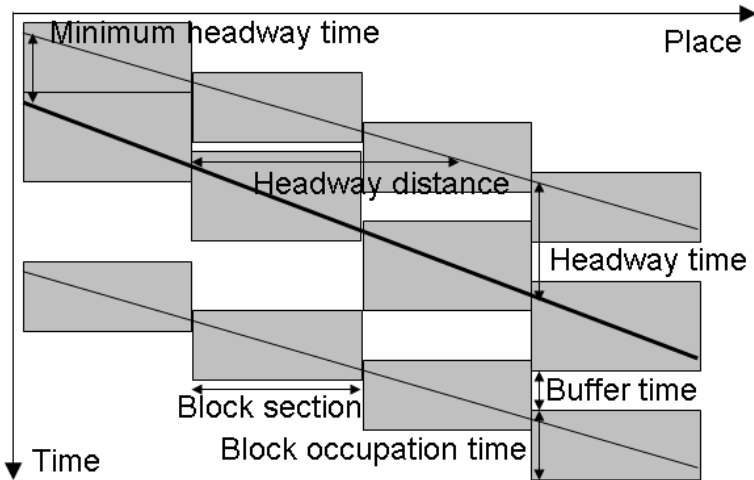


Figure 1: Time definitions [7].

3 The UIC 406 method

A detailed description of the UIC 406 method for capacity calculation is given in [9]. However, a brief description of the UIC 406 method will be given here.

Capacity consumption on railway lines depends on both the infrastructure and the timetable. Therefore, the capacity calculation according to the UIC 406 method is based on an actual timetable.

Timetables are created for the entire network and not only the line or line section which is of interest according to the capacity analysis. This means that the timetable in the analysis area depends on the infrastructure and timetable outside the analysis area – the so-called network effects [1][2]. These so-called network effects are not taken into account in the capacity analysis, why the capacity used according to the UIC 406 method will be less than or equal to the actual capacity consumption.

The capacity calculation is based on the compression of timetable graphs on a defined line or line section. All single train paths are pushed together to the minimum headway time, so that no buffer times are left. The compression of the timetable graph has to be done with respect to the train order and the running times. This means that neither the running times, running time supplement, dwell times or block occupation times are allowed to be changed. Furthermore, only scheduled overtakings and scheduled crossings are allowed.

4 The length of the railway section examined

The capacity calculation is based on compression of timetable graphs on defined line sections as described in section 3 and in [9]. However, the timetable graphs can only be compressed as the critical block section allows. Therefore, it is



important not to split the railway line into several smaller line sections uncritically, while the critical block section thereby might be excluded from the analysis.

The critical block section is the block section which is occupied for the longest time. For homogenous traffic the critical block section can be anywhere on the line, but normally the critical block section is located close to a station or halt due to the reduced speed [4]. For inhomogeneous traffic like the Coast Line (Kystbanen) in Denmark (cf. Figure 2) the critical block section is usually located where the fast trains catch up with the slower trains.

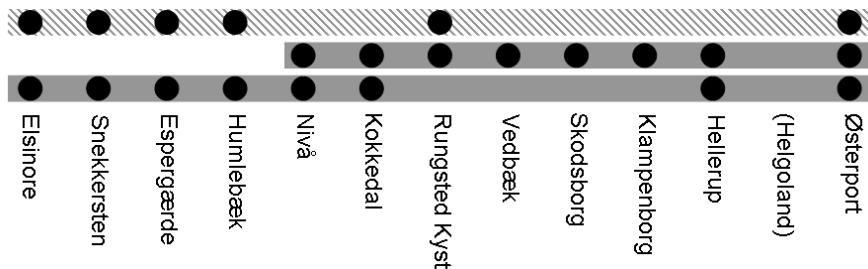


Figure 2: Train stopping patterns for the Coast Line in the rush hours (20 min service).

On the railway line between Copenhagen (Østerport) and Elsinore – the Coast Line – there is an inhomogeneous traffic (cf. Figure 2). The critical block sections or bottlenecks on the Coast Line are at Nivå and Hellerup where the trains catch up with each other. To analyse the capacity utilization of the bottlenecks it was decided to use the UIC 406 method, cf. Figure 3.

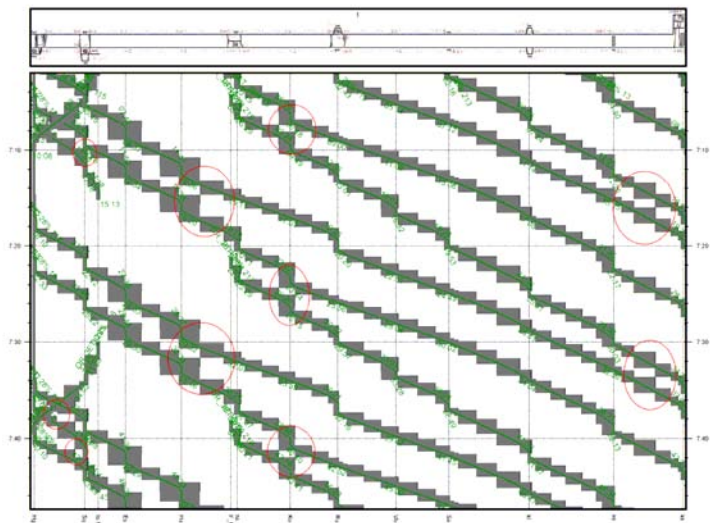


Figure 3: Timetable graph compression according to the UIC 406 method.

To test the UIC 406 method, 3 different line sections were examined. The 3 line sections were Kokkedal-Humlebæk, Helgoland-Klampenborg and the whole line section between Helgoland and Elsinore. The results from the analysis show a big difference in the capacity utilization at the bottlenecks, cf. Table 2.

Table 2: Capacity utilization on the Coast Line according to the UIC 406 method during the rush hour.

	Capacity utilization in %	
	Nivå	Hellerup
Short line sections	49	46
All line sections	94	84
Difference	45	38

Table 2 shows that the short line sections give much lower capacity consumption than the whole line section. The low capacity utilization at the short line sections can lead to the wrong conclusion that there is room enough to run more trains. However, running more trains will make it more or less impossible to keep a good punctuality (see section 6).

Based on the results from the Coast Line it can be concluded that it is important to examine the whole railway line and not just a smaller area when capacity analyses are carried out. However, it is not always possible to examine a whole railway line due to the analysis resources. Therefore, the effort has to focus on the analysis examining where the railway line can be divided into smaller line sections. Furthermore, it is necessary to be careful when comparing capacity utilizations and only compare relatively.

5 Changing between tracks at stations and at lines with more than two tracks

Compressing timetable graphs according to the UIC 406 method can lead to discussions on line sections with quadruple tracks. An example from the Capacity analysis of the line between Copenhagen and Ringsted [8] illustrates the problems. A freight train is running from Ringsted to the freight terminal at Høje Taastrup. Simultaneously with the freight train passing Roskilde, a regional train from Lejre towards Copenhagen leaves Roskilde, cf. Figure 4.

When compressing the timetable graphs there will inevitably be a conflict because the trains have to change use of track at either Roskilde or Høje Taastrup station. The capacity consumption of the line section depends on which train runs on which track between Roskilde and Høje Taastrup and thereby where the conflict between the trains occurs, cf. Figure 4.

Using the UIC 406 method on line sections with quadruple tracks it has (in Denmark) been decided to give priority to the track occupations of the actual timetable or a timetable with a minimum number of conflicts. In the Danish method it is only allowed to move one or more trains from one track to another if there is an unequal utilization of the tracks. Not until then is the consideration of passenger preferences taken into account. The passenger preferences are only

taken into account late in the process, since the UIC 406 method is used for capacity analysis.

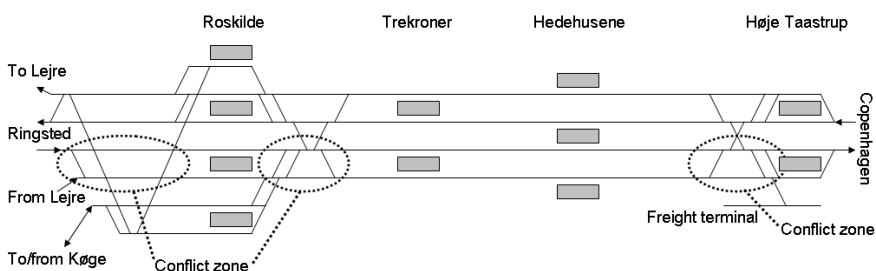


Figure 4: Conceptual track layout for the line section Høje Taastrup – Roskilde.

Quadruple track sections are usually used for fast trains to overtake slower trains. In the Danish way of using the UIC 406 method for capacity analysis, it has been decided that the order of the trains has to be the same as in the reference timetable (cf. Figure 5 a) in both ends of the line section when compressing the timetable graph (cf. Figure 5 b). The train order has to be the same in both ends of the line section even though the timetable graphs can be compacted more if the trains change the order (cf. Figure 5 c). The train order has to be the same due to the limitations of the infrastructure and timetable outside the analysis area – the so-called network effects [1]2].

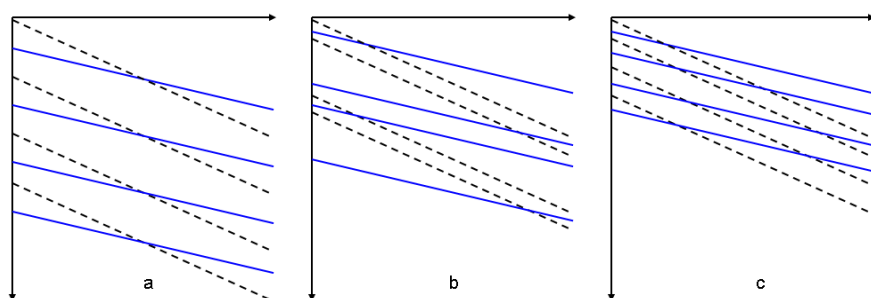


Figure 5: Timetable compression on quadruple track.

6 The possibility of using idle capacity to run more trains

The UIC 406 method describes the amount of capacity used on a certain railway line. However, not used capacity can not always be used to run more trains. In Figure 6 it is shown how the buffer times between the trains can give idle capacity by compressing the timetable graph and thereby make it possible to run an extra train.



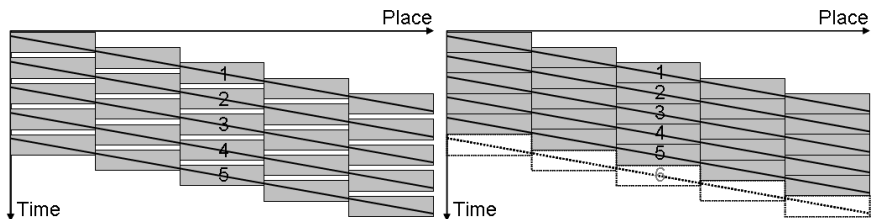


Figure 6: Usage of idle capacity.

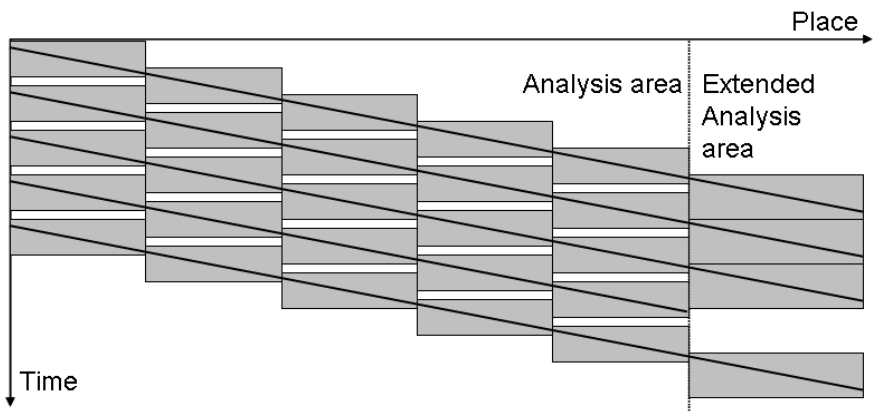


Figure 7: Limited possibility of compressing timetable graph.

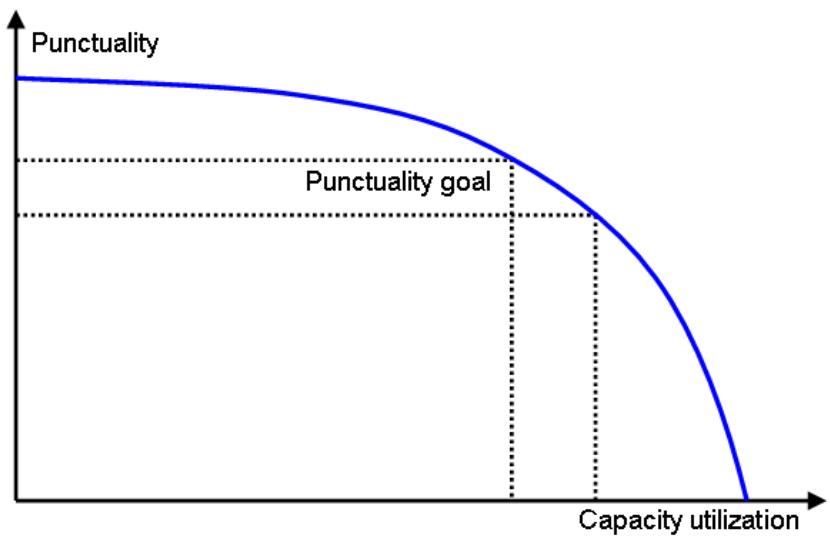


Figure 8: The coherence between punctuality and capacity utilization (based on [5]).



It is, however, not always possible or wise to use the idle capacity or buffer time to run more trains. If there is a larger block section outside the evaluation area, it is not possible to run an extra train due to the lack of capacity outside the evaluation area, cf. Figure 7.

Even though it is possible to run more trains, it is not always wise to do so because it will reduce the buffer times. By reducing the buffer times, the risks of secondary delays are increased. Furthermore, the dispatching of the trains is made more difficult due to more trains in the system. Everything considered, the idle capacity can not always be used to run more trains. A weighing of the punctuality of the trains and the capacity utilization has to be done, so that the punctuality will not drop below a certain limit, cf. Figure 8.

7 The possibility of using the plan of operation instead of the exact timetable

It is difficult to analyse the capacity of a railway line which has not yet been built since the timetable for the opening year is unknown. Furthermore, the timetable can be changed over time. Therefore, examining the capacity utilization based on the plan of operation instead of the final timetable for the opening year is preferred. Since the capacity utilization is influenced by the order of the trains, it is necessary to make some assumptions of the train order or use methods used for successive calculation.

Using successive calculation, the average capacity utilization can be calculated as a weighted average of one (or more) likely value(s), the maximum value and the minimum value as shown in eqn (1):

$$\text{Average} = \frac{\text{MinValue} + (3 \cdot \text{ActualValue}) + \text{MaxValue}}{5}. \quad (1)$$

The minimum value (MinValue) and the maximum value (MaxValue) of a plan of operation can be found by arranging the trains, so that the trains utilize as little or much capacity as possible, cf. Figure 9.

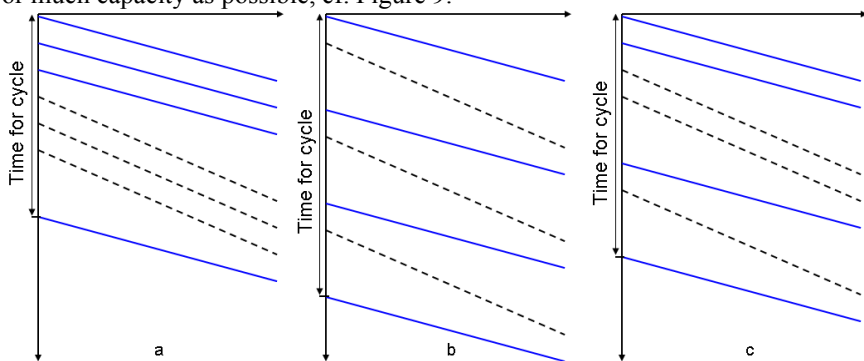


Figure 9: Plan of operation utilizing as little (a), as much (b) capacity as possible and an “actual” capacity consumption (c).



Beside the values for the minimum and maximum capacity utilization it is also necessary to calculate 3 actual values for the capacity utilization, cf. eqn (1). This is due to the fact that the infrastructure outside the section of examination can limit the number of ways the trains can be ordered.

8 Conclusion and perspectives

The analysis has shown that the capacity utilization on railway lines is very responsive to the network examined. Therefore, the capacity utilization should only be compared relatively. In Denmark it has not yet been decided where to split the railway lines into smaller line sections – this work still remains to be done.

When there is a quadruple track available, it has been decided that the track occupations of the actual timetable should be used. If there is no actual timetable the timetable with the minimum number of conflicts should be examined instead. It is furthermore only allowed to move a train from one track to another if there is an unequal utilization of the tracks. The conditions for passengers transferring to e.g. busses are not taken into account.

Even though the capacity analysis shows that it is possible to run more trains in the section analysed, it is not always possible. The analysed line section can be too short to see that it is not possible to run more trains (e.g. due to capacity restrictions outside the analysis area) – the so-called network effects.

All the analyses have been carried out in the timetable and simulation software RailSys, but can also be carried out in the timetabling system STRAX/TPS which is used by the Danish railway agencies [6]. The result of the analysis is a common Danish method to evaluate the capacity utilization of railway lines in Denmark.

References

- [1] Hansen, S., Landex, A. & A.H. Kaas, The Network effects of Railway Investments, *Proc. of the 10th International Conference on Computers in Railways*, eds. J. Allan, C.A. Brebbia, A.F. Rumsey, G. Sciutto, S. Sone & C.J. Goodman, 2006
- [2] Hansen, S., Large Transport Infrastructure Investments and their Strategic Impacts with a Special Focus on Enterprises, *PhD. thesis at Centre for Traffic and Transport, Technical University of Denmark*, 2004 (in Danish).
- [3] Höllmüller, J. & Klahn, V., Implementation of the UIC 406 capacity calculation at Austrian railways (ÖBB). *Proc. of the 1st International Seminar on Railway Operations Modelling and Analysis*, eds. I.A. Hansen, F.M. Dekking, R.M.P. Goverde, B. Hindergott, L.E. Meester, The Netherlands, 2005
- [4] Kaas, A. H., Methods to calculate capacity of railways, *PhD. thesis at Department of Planning, Technical University of Denmark*, 1998 (in Danish).



- [5] Kaas, A. H., Punctuality model for railways. *Proc. of the 7th International Conference on Computers in Railways*, eds. J. Allan, R. J. Hill, C. A. Brebbia, G. Sciutto & S. Sone, pp. 853-860, 2000
- [6] Kaas, A. H. & Goossmann, R., Implementation of the Timetable Planning System STRAX/TPS in Denmark. *Proc. of the 9th International Conference on Computers in railways*, eds. J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto & S. Sone, pp. 93-102, 2004
- [7] Landex, A. & Kaas, A. H., Planning the most suitable travel speed for high frequency railway lines. *Proc. Of the 1st International Seminar on Railway Operations Modelling and Analysis*, eds. I.A. Hansen, F.M. Dekking, R.M.P. Goverde, B. Hindergott, L.E. Meester, The Netherlands, 2005
- [8] Strategic analysis of Copenhagen-Ringsted – Comparing solutions for extending the railway capacity; The Danish Rail Authority, October 2005, Online. http://www.trafikstyrelsen.dk/graphics/Synkron-Library/Trafikstyrelsen/Jernbane/Strategi/strategi_beslutningsgrundlag_.pdf (in Danish).
- [9] UIC leaflet 406, Capacity, 2004, UIC International Union of Railways, France
- [10] Wahlborg, M., Banverket experience of capacity calculations according to the UIC capacity leaflet. *Proc. of the 9th International Conference on Computers in railways*, eds. C.A.- J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto & S. Sone, pp. 665-673, 2004



A method to estimate passenger flow with stored data at ticket gates

S. Myojo

Railway Technical Research Institute, Japan

Abstract

In order to improve railway diagrams, it is necessary to grasp the flow of railway passengers in detail. However, it is impossible to accomplish this goal only by counting the number of on-board passengers. We have developed a method to estimate the passenger flow based on OD (Origin & Destination) matrix data collected from ticket gates in hundreds of stations and timetables of thousands of trains. By analyzing the estimated results, it is possible to obtain the histograms of the number of passengers aboard each train for each destination, the number of passengers boarding and disembarking at each station, the number of transfer passengers at each station and other relevant factors. We have also developed a proto-type system based on this method assuming that all passengers choose shorter time paths. The estimated number of on-board passengers with this proto-type system has a strong correlation with the number reported by train conductors.

Keywords: *passenger flow, ticket gate, OD.*

1 Introduction

In order to improve railway diagrams or passenger announcement, it is necessary to grasp railway passenger flow in detail including the number of on-down passengers on each train and on-board number per each destination. In Japan, the Ministry of Land, Infrastructure, and Transportation holds the transportation census in order to perceive a passenger flow of public transportation in three major metropolitan areas. Japanese government utilise the census data for planning railway network and public transport operators do for scheduling trains or planning capital investment. The census provides very detailed data of passenger flow because passengers corporate with the survey fill their choice of



transportation mode on the census day. However, there is a problem that Japanese government holds it only once every five years because it is a huge-scale survey.

On the other hand, rolling stock incorporated recently is equipped with a device to measure the total weight of on-board passengers and calculate the congestion rate. This function will make it possible to obtain a total of passengers aboard all trains. However, it is time-consuming until every car is equipped with such function, and this function does not help to obtain on-down number and passengers' destination. Therefore, it is necessary to realize a method to grasp on-board number of each train on any day under the condition where not every car is equipped with that function and/or to grasp as detailed data as the Transportation Census.

We have developed a method to estimate passenger flow based on OD matrix data stored by ticket gates at stations. Although OD data do not provide directly any information about lines and trains whom passengers took, the method speculates their choice of paths including their lines and trains. By analyzing estimated result, it is possible to obtain the number of passengers aboard each train for each destination, the number of passengers boarding and disembarking at each station, the number of transfer passengers at each station and other relevant factors.

We also have developed a proto-type system based on this method assuming that all passengers select shorter time paths. It takes approximately one hour for this system to estimate one day's passenger flow of a railway network that contains more than 300 stations, and the estimated numbers of on-board passengers with this proto-type system has a close correlation with the number of passengers reported by conductors.

2 OD data during each period

OD signifies Origin and Destination. The term "OD data" means each number of passengers of each OD pair (Figure 1). Ticket gates count the number of passengers of each OD pair in each partitioned period of one day.

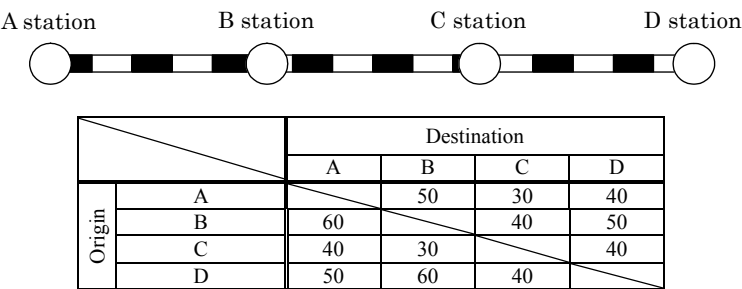


Figure 1: An example of OD data.

OD data has the following advantages:

- It is possible to obtain the data on any date because gate machines work everyday.
- The data is almost complete enumeration because gate machines count almost all passengers.

In spite of such advantages, OD data has a disadvantage that it does not provide directly any information about lines and trains that passengers took. Therefore, we have developed an estimation method assuming their routes and trains.

3 Proposed method to estimate a passenger flow

3.1 Basic idea of the method

The proposed method estimates passenger flow with following steps.

- (1) Select a pair of OD from OD data gathered from ticket gates.
- (2) Search all paths that those who travel the OD selected in procedure (1).
- (3) Calculate the number of passengers on the paths searched in procedure (2) during each period.
- (4) Store the number of passengers calculated in procedure (3).
- (5) Repeat procedures (1), (2), (3) and (4) until all pairs have been finished.
- (6) Calculate the total number of passengers that pass each link.

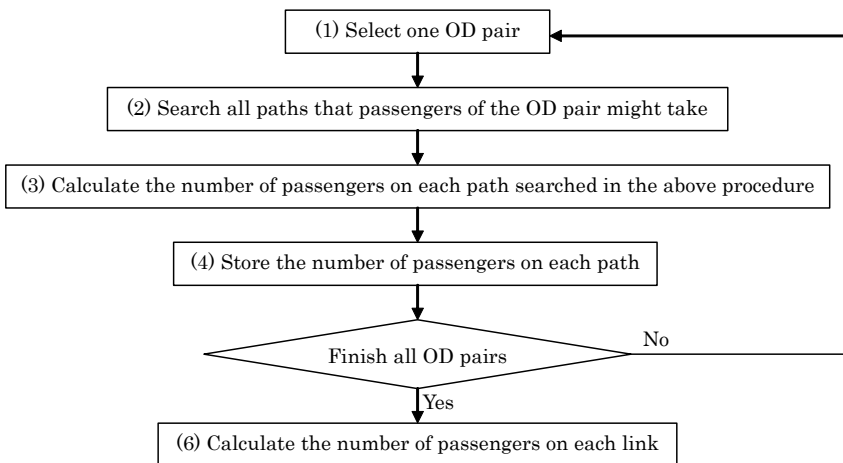


Figure 2: Procedures to estimate passenger flow.

3.2 Procedure to search paths

As described above, it is necessary to determine passengers' choice of their paths in order to estimate a passenger flow using OD data during each period. The paths in this method contain passengers' trains as well as their train lines. In



other words, a path not containing a different train from another one is not same with the one though they contain same railway line(s) at all. Therefore, estimated result can provide detailed data about passenger flow.

It is possible to search paths containing trains by transposing behaviours such as taking a train, disembarking, change of trains and other action to a directed graph. Figure 3 shows an example of components that make up a directed graph. In this graph, there are nodes representing departure from a station on a train and arrival at a station, and links representing travelling on a train, waiting for departure and change of trains. The weight (or length) of each link is time spent for the behaviour of travelling aboard a train, waiting for departure and change of trains when searching shorter time paths. When searching paths with fewer changes of trains or containing train(s) of fewer congestion rates, we add transfer resistance and congestion rate to each link in terms of time.

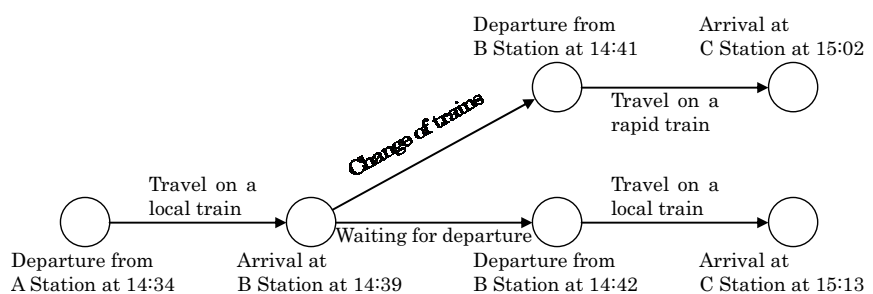


Figure 3: Components to make up a directed graph for path finding.

3.3 Calculation of the number of passengers on the searched paths

In this procedure, the estimation method distributes the number of passengers who travelled the selected OD pair to searched paths during each period. Equal allocation is the simplest method. With this method, the number of passenger on each path is $\frac{Q_{ij}}{n}$ where Q_{ij} is the number of passengers of the OD pair i during the period j and n is the number of paths. When difference between paths is not small, equal distribution is not suitable but logit model is. With distribution method using a logit model, the estimation method calculates the number q_k of passengers on the path k with the formula (1) where V_k is the utility value of the path k .

$$q_k = \frac{e^{V_k}}{\sum_{l=1}^n e^{V_l}} \cdot Q_{ij} \tag{1}$$

3.4 Calculation of the number of passengers on each link

The estimation method calculates the number x_i of passengers on the link i with the formula (2) where q_{ij} is the number of passenger of the path j who moved on the link i and M is the number of paths of all OD pairs. The number q_{ij} is equal to the number of passengers of the path j if the path j contains the link i , and the number q_{ij} is zero if the path j does not contain the link i .

$$x_i = \sum_{j=1}^M q_{ij} \quad (2)$$

The whole of x_i are the result of estimation.

4 Various data obtained from the estimated result

4.1 A proto-type system

Based on the method, we have developed the simplest proto-type system that searches shorter time paths and allocates the number of passengers equally. We have applied this system to estimate passenger flow in the real railway network in the following condition:

Number of stations: 392

Number of trains: 4,785

Number of OD pairs: 30,604

The specification of the directed graph to search paths in this railway network is below:

Number of nodes: 216,183

Number of links: 506,233

The system can estimate within an hour with PC whose CPU is 3.6GHz Pentium® 4 processor.

4.2 Number of on-board passengers

By sorting the numbers of the passengers on links, as shown by thick arrows in Figure 4, contained in a train into near order from the starting station of the train, it is possible to obtain the change of number of on-board passengers of the train. Figure 5 shows an example of a histogram of the estimated number of passengers of a real train. The train started from (A) station and stopped at every station from (B) to (J). The shaded bars represent the number of passengers while travelling and white bars those while waiting.

4.3 Number of passengers boarding and alighting at each stop

By comparing the number of passengers before a stop and at the stop, it is possible to obtain the number of passengers who alighted at the stop.



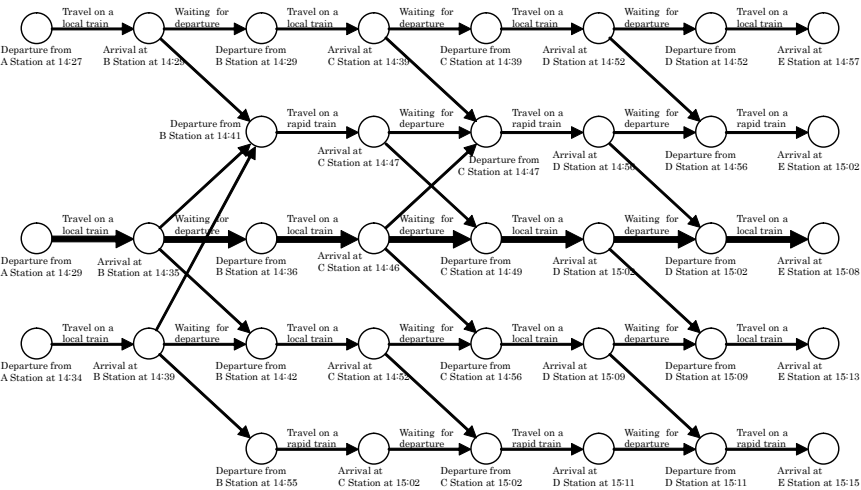


Figure 4: Links that are part of a train.

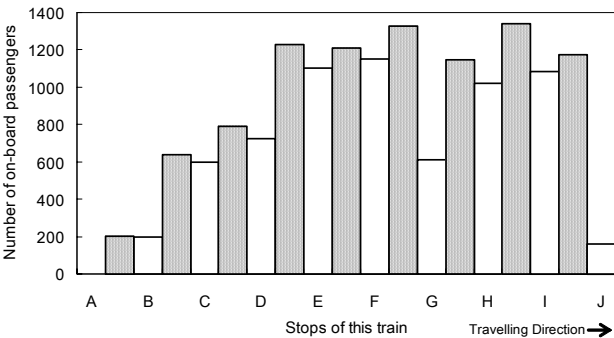


Figure 5: Numbers of on-board passengers.

By comparing that before departure and after departure, it is also possible to obtain the number of passengers who boarded at the stop. Moreover, it is possible to distinguish the number of passengers transferring to other trains from that of all passengers who alighted at a stop because the result of estimation provides the number of passengers on links representing transferring to and from the train shown by double-lined arrows in Figure 6. It is also possible to distinguish the number of passengers transferring from other trains from that of passengers who boarded on the train.

Figure 7 shows an example of a train. Bars of positive value represent the number of passengers who boarded train, and those of negative value the number of alighting passengers. It is very difficult to obtain such data by only counting the number of on-board passengers.

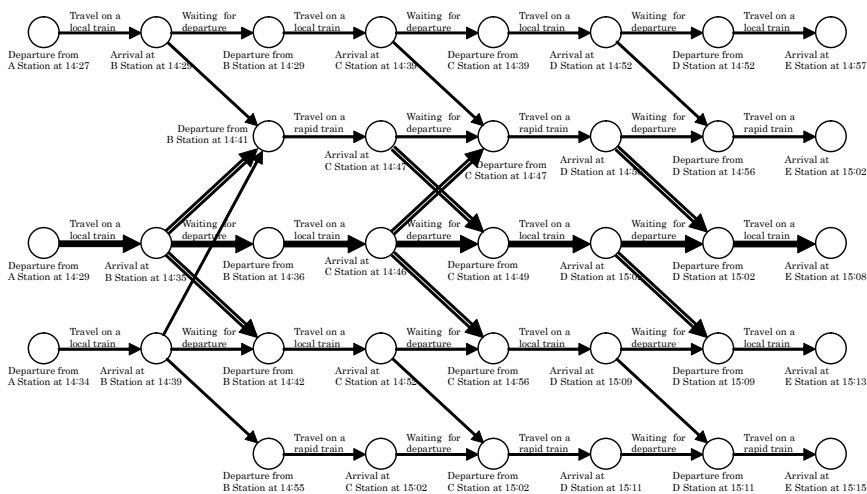


Figure 6: Links representing transfers.

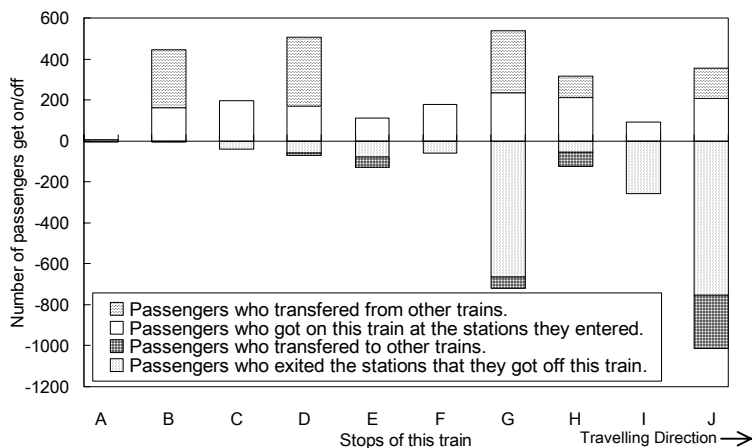


Figure 7: Number of passengers boarding and disembarking at each stop.

4.4 Number of on-board passengers for each destination

By integration of the number of passengers on each link contained in paths concerning with specific destinations, it is possible to obtain the passenger flow data about specified destination. That is, it is possible to obtain the number of on-board passenger for every destination. Figure 8 shows an example.



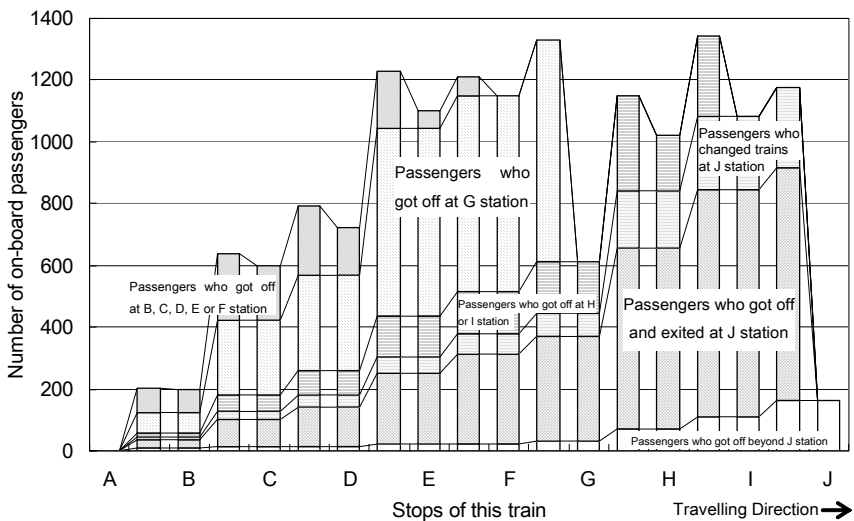


Figure 8: Number of on-board passengers for each destination.

4.5 Traffic value between two stations

Figure 9 indicates the number of passengers on links between two stations as compiled with thick arrows. It is possible to obtain a change in traffic value between the stations, figure 10.

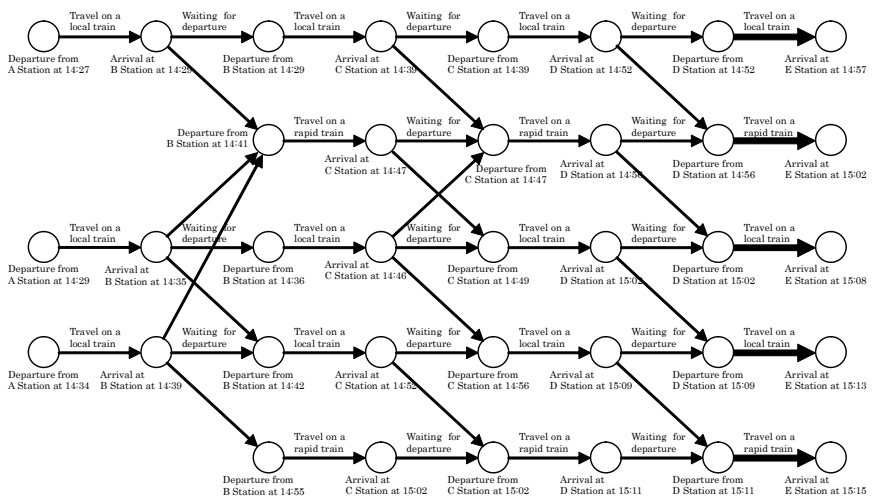


Figure 9: Links between two stations.

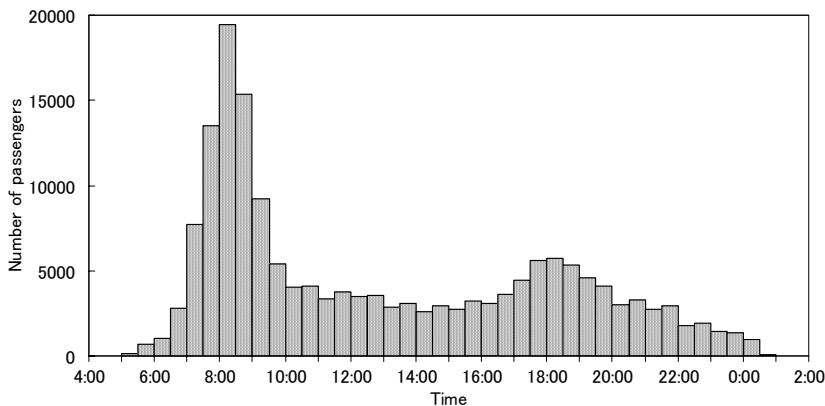


Figure 10: Histogram of traffic value between two stations.

5 Verification of estimated result

By comparing the reported number of passengers by train conductors, we have verified the number of on-board passengers estimated by the proto-type system. Train conductors of nearly a half of all trains report the maximum number aboard between two or more stations by visual counting. Figure 11 and 12 show the comparative results. Figure 11 shows the result of comparison on a line where only local trains run. Figure 12 shows that on another line where rapid trains make connections with local trains. In both lines, especially on a line of local trains only, correlation coefficient is high.

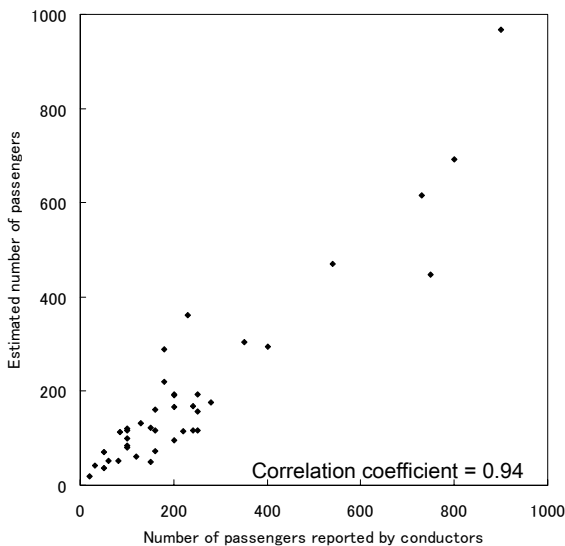


Figure 11: Comparison on a line where only local trains run.



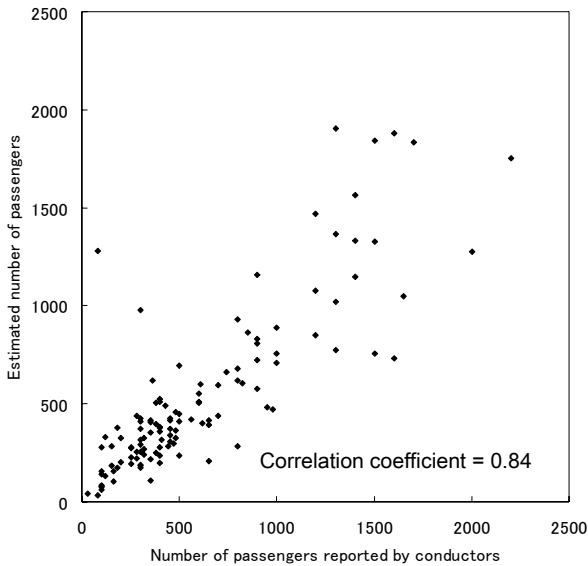


Figure 12: Comparison on a line where rapid trains make connection with local trains.

6 Conclusion

It is possible to estimate the passenger flow and to obtain detailed data containing the number of boarding passengers and disembarking passengers, the number of on-board passengers for each destination using OD data during each period stored in ticket gates. Moreover, the number of on-board passengers estimated by the simplest proto-type system has a very close correlation with the reported number of passengers by train conductors. In a line where only local trains are under operations, the simplest system can be practical.

References

- [1] Suzuki S, Goto K, Estimation of Passenger Flow from Vehicle Air Spring Pressure and Automatic Ticket Checking Data, RTRI Report No.7, pp. 39 44, 2004 (in Japanese)
- [2] Ieda H, Akamatsu T, Takagi J & Hatanaka H, Commuter's Benefit Evaluation of Train Scheduling by Network User Equilibrium Model, Infrastructure Planning Review, pp. 177 184, 1988 (in Japanese)
- [3] Nagasaki Y, Asuka M & Koyama K, Fast Estimation Method of Railway Passengers Flow, 2005 (in Japanese)



Analysis and optimisation of railway nodes using simulation techniques

A. Kavička¹, V. Klima² & N. Adamko²

¹*Jan Perner Transport Faculty, University of Pardubice, Czech Republic*

²*Faculty of Management Science and Informatics, University of Žilina, Slovak Republic*

Abstract

Railway nodes represent sensitive points of railway networks, the efficiency of which influences the effectiveness of network processes. Those nodes belong to the most complex service systems involving sophisticated technological processes and are equipped with quite complicated and costly technical devices. Design of those systems and the organization of their operation have to facilitate the required capacity together with a high quality of service processes as well as minimal costs of resources. That intention can be achieved using various measures, e.g. track infrastructure modification, utilization of alternative kinds of resources, improvement of scheduling and rostering, enhancement of technological processes, adjustment of decision making strategies or entire system reengineering (that approach can be applied to a design of a new node). The application of exact mathematical optimisation methods is quite limited because of the high complexity of the above-mentioned systems. Computer simulation offers a quite flexible and credible technique eliminating the disadvantages of precedent approaches, i.e. a simulation model of railway node (substituting an existing or planned system) enables one to investigate all mentioned ways of an increase of node capacity. Efficient methodologies are introduced aimed at the solution of capacity problems related to railway nodes using the simulation software tool Villon, which was originally determined for simulations of railway marshalling yards; however its flexible hierarchical software architecture enabled us to extend its scope to other kinds of transportation nodes (including e.g. automobile traffic).

Keywords: transportation node, simulation model, planning, optimisation.



1 Transportation terminals

The most important and evidently the most costly part of a logistic chain is the transfer process. This part is often called transport logistics. The transfer process itself is composed of a movement along the transportation route and of necessary manipulations (service processes) applied to transportation means and to transported commodities at specialised locations called (transport) *terminals*. At the terminals, individual entities may originate, disappear, and be subject to service or to transformation. Several kinds of elements are distinguished in the terminal structure:

- Terminal *infrastructure* (transportation paths, security and information elements, buildings and other static equipment, e.g. platforms, parking lots, etc.),
- *Mobile service resources* (human resources, transport and manipulation devices, etc.),
- *Elements subject to a service* (cars, railway carriages, containers, passengers, etc.) and
- *Control system for service processes*.

Undoubtedly, the most complicated situation from the point of view of manipulation and process control occurs in terminals, where the infrastructure is composed mostly of set of tracks, where the mobile service resources are mostly locomotives and where the elements under service are also mostly railway vehicles. Here belong railway logistic terminals and terminals with a significant portion of a railway infrastructure. Exactly the terminals with these characteristics are in the focus of our attention.

Let us see examples of the railway logistic terminals, where the common problems concerning the infrastructure design and planning and the traffic control appear, and where also the similar approach to their solution is applicable:

- *Marshalling yards* contain usually only the trackage and the prevalent activities are transformation operations with the rail vehicles (coupling, uncoupling, sorting, various kinds of shunting, etc.).
- *Passenger railway stations* are also equipped with trackage, but the prevalent activities are the service operations (entering and leaving the carriages, cleaning, etc.). The elements subject to a service are passenger carriages and also passengers in the entrance hall.
- *Industrial sidings* are often combined with road transport and so they contain infrastructure, service resources and elements subject to service of both types. In heavy industry plants, the set of tracks may be extremely large and may even contain an independent train forming part. Here, complex operations are characteristic, namely complicated manipulations and also loading and unloading of different substances.
- *Specialized railway logistic centres* serve usually for maintenance, reparations, cleaning and other service and for completion of train sets (railway depots).



- *Terminals for multimodal transport* (e.g. container terminals) contain also a heterogeneous infrastructure and various elements (railway, road, water transport elements).
- *Transport nodes*, which, from the point of view of infrastructure problems and traffic operation problems, should be understood in a broader sense as an integration of several types of terminals. For example, a need may arise to integrate goods railway station, passenger railway station and a corresponding depot into a single transport node.

2 Problems of infrastructure planning

Generally, regardless the terminal type, the problem of infrastructure planning (e.g. trackage design) inside the terminal presents the task to define elements of the infrastructure and the task to integrate them into a system. The infrastructure planning need arises either in the realisation of long-term strategic aims or in the optimisation of the operation of standing terminals.

Among *strategic* aims are namely:

- Decision to construct a new terminal or a new part of a terminal.
- Technical reconstruction of a terminal infrastructure, usually followed by modernisation of its elements.
- Enlargement of infrastructure with the aim to increase its capacity, if the service increase is expected.
- Decision to reduce the active infrastructure if the service depression is expected.
- Decision to concentrate investments into modern and well equipped large terminals and to suppress operations in other terminals.

Among *operational* problems in infrastructure planning are:

- Terminal operation is too costly or terminal is working below required capacity while solutions via optimisation of mobile resources and via optimisation of technological processes are exhausted.
- It is evident that the infrastructure is not correctly balanced, i.e. some parts are insufficiently dimensioned and are causing operational bottlenecks while other are over dimensioned and are not sufficiently exploited.

The problem of infrastructure planning is in other words a problem of optimal configuration of the terminal infrastructure. A correct solution of an optimisation problem requires, of course, definition of an optimisation criterion. But in this case it is not easy at all and we even do not intend to try to do it. Instead we will point out some objectives we are trying to achieve, which are often contradictory:

- a) Minimal infrastructure size is desirable because of significant construction, maintenance and operational costs.
- b) Sufficient size of infrastructure must be defined to cope satisfactorily with a stochastic nature of required service volume.



- c) Configuration of infrastructure elements should be balanced to avoid bottlenecks on one side and idle parts on the other.
- d) Infrastructure configuration should have “reasonable” reserves of capacity to cover long term or short term increase of service requirements and random jumps.

3 Problem of optimal infrastructure configuration

Any solution of optimisation of infrastructure configuration should observe the following facts:

- Realisation of an accepted solution is typically extremely costly and thus the solution should be a long term one. It is practically impossible to make additional corrections in the case of a wrong decision.
- Infrastructure is only a resource, which serves the goal – traffic operation. *Configuration of an infrastructure is satisfactory only if the operation on it is effective.* Therefore, if we want to judge the suitability of an infrastructure, we have to know and to describe in detail (make a model) the traffic realized on it.
- The traffic operation of a terminal is an extremely complex system. It is a dynamic system with complex relations among its elements where the service processes are mutually intertwined mostly with stochastic interactions.

The question is what can be done from the management side to ensure the most objective decision and to avoid an erroneous decision.

Due to the complexity of the system and its stochastic behaviour, the use of exact mathematical methods is significantly limited. Classic planning methods, on the other side, use simplified models of traffic operation which do not take into account the stochastic features of the system and do not involve detailed behaviour of mutually dependent dynamical technological processes. Instead, mean and normative values are used together with expert knowledge based on past experience. Frequently, the result is a solution which, confronted with a real traffic, appears to be unsuitable.

Where can we find a way out of this situation? Exact mathematical solution, as we can see, is not practicable. The only feasible way is to accept a good suboptimal solution and to find it in an environment, which is capable to reflect satisfactorily the complexness of the infrastructure and of the traffic and at the same time it is capable to produce understandable solutions based on verified arguments. Experimental environment which satisfies these conditions is a *simulation model of a terminal*, embodied in hardware and software form as a substitute for a real or designed terminal, which respects exactly the infrastructure composition and describes in detail all traffic processes in their complexity. Simulation model is used as an experimental environment with the objective to evaluate all consequences of proposed infrastructure configuration. Evidently, the effectiveness of the traffic is a criterion of the evaluation.



4 Simulation techniques

Simulation of systems is a method supporting analysis, design and optimisation of real systems. Three steps can be distinguished:

1. Substitution of a real system by its simulation model.
2. Experimenting with the simulation model with the aim to disclose its properties, behaviour and reactions under selected conditions.
3. Application of obtained results in the real system (existing or designed)

Simulation model should be as accurate and detailed as possible so that its results are applicable to the reality. At the other side, there is also certain limit of detail, which should not be trespassed.

The software product Villon belongs to simulation tools which are based on above described principles and was successfully verified in practice. Villon is not only a very efficient tool for verification of infrastructure modifications but it is also very suitable mean for investigation of possibilities how to increase the efficiency of service resources and improve decision activities. Villon enables the user to construct complex, detailed and interactive simulation models, make experiments with them and analyse the results. Simulation model developed in the frame of experimental environment of the tool Villon can be an effective aid for solution of above described problems of infrastructure planning, both strategic and operational.

The tool Villon itself, however, does not give automatically ready-to-use optimal solutions of complex problems but offers an experimental environment, a “laboratory” for investigation of various configurations of a terminal infrastructure configuration and all consequences of their application. Simply said, using Villon, a user-experimenter can answer questions “*What happens, if...?*”. It is supposed that the user himself is expert in technology and, moreover, cooperates closely with the management officers of the terminal.

Several basic stages of work with Villon can be distinguished:

- a) *Acquisition, processing and analysis* of data concerning real or planned terminal.
- b) *Construction of infrastructure model* (object of investigation).
- c) *Construction of dynamical operation (traffic) model*, which serves for investigation of the terminal properties with proposed infrastructure configuration.
- d) *Experimenting* with the simulation model. It is an iterative process of consequent runs of *simulation experiments* with modified parameters with the objective to disclose the behaviour of the system in correspondence with the given set of parameters leading to the solution of the problem. It is evident that a number of simulation experiments should be performed. This number depends on the degree of expertise of experimenter, on the quality of the simulation model and also on tools available for evaluation of experiments. A correct evaluation of results undoubtedly shows the direction and next steps in the quest for solution.
- e) *Analysis of results* of simulation experiments. Villon enables the user to follow on-line animation of movements of all mobile resources and of



technological activities. Moreover, Villon is capable to submit information on properties of all both static and dynamic terminal elements including clients of the system, via so called Explorer. The broad scope of tools includes the possibility to obtain up-to-date statistics about the performance (state) of all mobile and fixed elements of the terminal. This information is given in a graphic form.

It is clear that during the simulation run we are not able to discover all aspects of the system behaviour. Villon however offers the possibility of recording simulation run into a simulation protocol which can be used later as a data source for post-simulation evaluation of results. Again, Villon offers a user a set of tools for statistical evaluation of protocol data, for construction of schedules of resources activities in graphic form, exploitation of infrastructure, etc. If needed, all recorded data can be exported e.g. to MS Excel format.

Credibility of the experiment results strongly depends on the good fit of the infrastructure model and of the traffic operation model to the reality. In the following two sections these models and their properties are dealt with more closely.

5 Model of infrastructure

Steps leading to the construction of an infrastructure model are best explained on the example of building the model with the most complicated infrastructure, which is a railway (track) infrastructure. The infrastructure model (fig. 1) is created by transformation of a map (plan) which is available either on paper or in electronic form. The obtained infrastructure model guarantees an absolute accuracy of the fit. Thanks to it, an accurate, not only simplified model of trackage is used in the simulation and therefore we can evade distortions caused by rough approximations of distances and their influence on movements of mobile elements in the simulation model.

Transformation of original plans constitute so called *physical level* of an infrastructure model. It means that a trackage composed of basic elements, tracks, simple switches, cross switches, track crossings, is available to the model builder (tracks can be divided into sections corresponding to the trackage organisation or into isolated sections for the signalling system). In this level, however, professions of individual tracks are not yet defined. The notion "profession" means the purpose for which the tracks serve in the trackage, e.g. arrival of terminating trains, connecting tracks to depots for uncoupled train locomotives, hump tracks, sorting tracks, etc. Information on the track profession is important for the model building in the phase of technology definition and definition of technological activities of the service of train sets. For example, in the technological activity "Transfer of shunting locomotive to the train set", profession of track, used for this transfer, must be defined. Professions are assigned to individual track in accordance with the track professions in the trackage of the real infrastructure. Assignments of professions constitute a *logical level* of the infrastructure model.



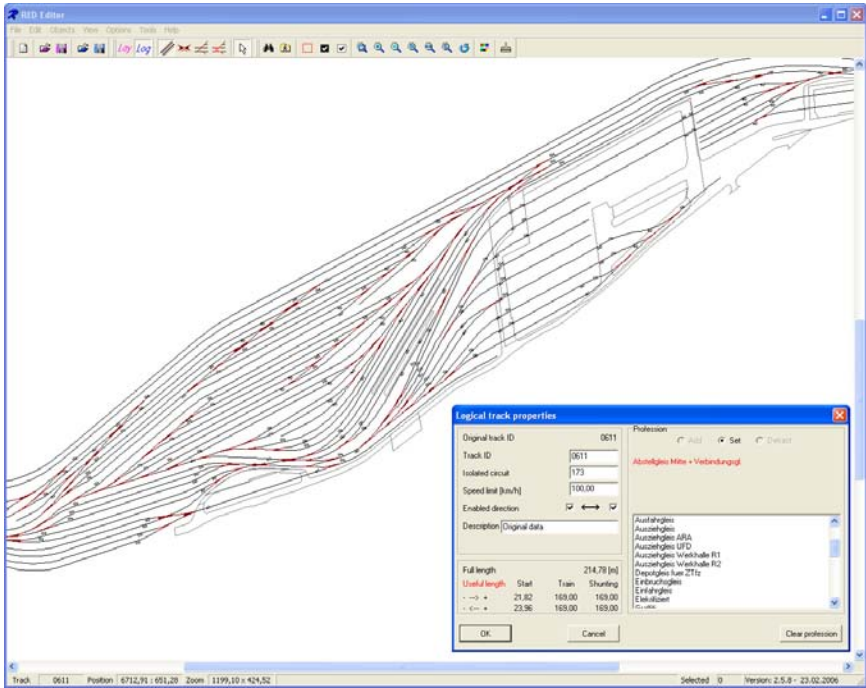


Figure 1: User interface of the programme RED presenting a detail depot infrastructure and a track properties dialog box.

After physical level of the infrastructure model, list of track professions and logical level of the infrastructure are defined, routes for the mobile elements are created. Moreover, also dynamically ad hoc created routes may be used, which are found automatically during the simulation experiment, where an actual availability of tracks is considered.

6 Traffic operation model

To verify the suitability of a proposed configuration of infrastructure in a real terminal, its operational verification is important. It is done by a series of simulation experiments with various sets of traffic conditions. These simulation experiments may reveal eventual flaws in the infrastructure configuration, which may cause, that expected performance of the terminal or the smooth run of local technological processes are not reached. Then, another series of experiments is performed with modified configuration aimed at elimination of ineffectiveness. In an extreme case even the misconception of the whole infrastructure can be proved by the simulation experiments and its reconstruction is necessary. Of course, this approach is very valuable for eventual investors, who will decide about their investments in the infrastructure construction on the basis of successful simulation verification. Once more we stress that a good fit to reality



must exist, all important factors concerning real traffic operation must be observed. In the simulation tool Villon, dynamic model of the terminal traffic operation consists of the following components:

- a) *Mobile service resources*, which include individual *workers*, operating in the terminal field (e.g. shunters, couplers, external transiteurs, locomotive guards, etc.), *locomotives* (e.g. shunting, hump or train locomotives, etc.).
- b) *Technological processes* formally describing standard technological processes, including requirements for assignment of necessary resources. Good examples of a suitable and clear formal presentation of technological processes are so called networks (directed acyclic graphs).
- c) *Objects under service*, describing temporary elements subject to various type of service is given (e.g. trains or wagons, trucks, containers, etc.).
- d) *Operations control*, focused namely to (i) setting priorities in processing of individual objects under service, (ii) management of all types of service resources, i.e. of infrastructure and of mobile elements which are assigned to or taken away from clients – service processes, with respect to given rules.

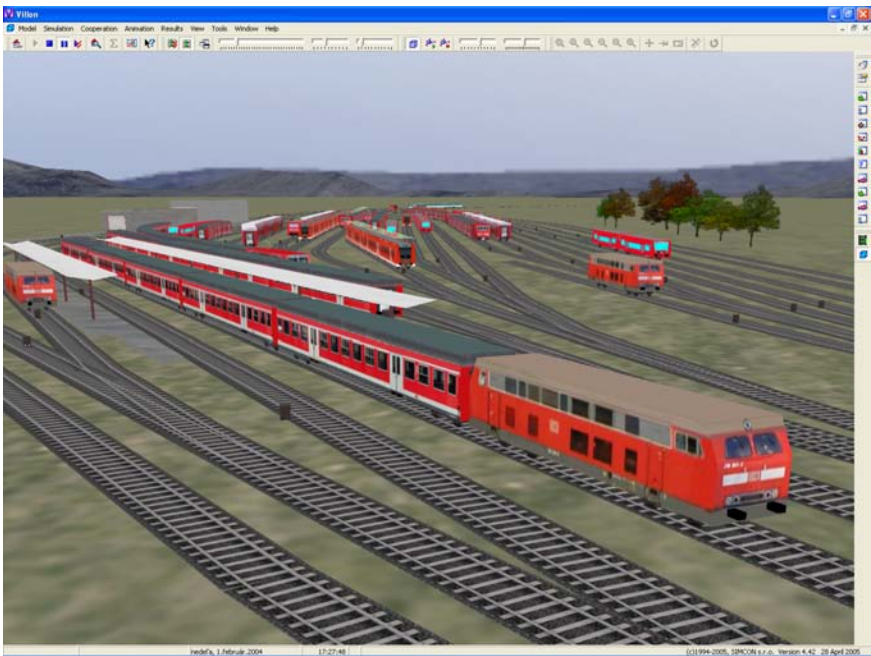


Figure 2: Animation of vehicle movements.

Experimental simulation environment of the tool Villon illustrates the operation of investigated traffic using an on-line animation (fig. 2), which enables a detailed following of its development during the working period and offers a broad scope of post simulation statistical evaluations of the experiment results.



7 Applications

In section 2, typical problems in the infrastructure planning were presented, where simulation technique may be an efficient tool for their solution. Simulation tool Villon assisted successfully to the solution of most of these problems, always with a significant economic profit.

Villon was used for example for the infrastructure planning and verification of traffic operation in the train set depot in Ulm, Germany. Here, the problem of trackage configuration was complicated by problems of location and configuration of individual service modules of the depot (internal and external cleaning station of train sets, fuel supply device, repair and maintenance facilities, etc.). Another example of the Villon successful application in planning a new terminal is Slovak marshalling yard Teplička nad Váhom near Žilina, whose importance at present is increasing substantially.

Changes in the organisation of train forming system in Austrian and Swiss railways led to cancelling of several small marshalling yards and modernisation and extension of greater and more efficient ones. The customer flows originally processed in the cancelled stations had to be redirected into the remaining yards. Villon was used for verification of suitability and efficiency of new infrastructure solutions in Linz (Austria) [9], Basel and Lausanne (Switzerland).

Attention was also paid to infrastructural and technological changes connected with the traffic operation of passenger railway stations [2]. Last year, Villon was applied with success in the investigation of planned traffic operations in Prague, Masaryk railway station, in Brno (design of a new railway station), both in Czech Republic, and in one passenger railway station belonging to Beijing railway node in the People's Republic of China.

Industrial railways must be flexible enough to be capable to cope with an increase or decrease of company production, which immediately reflects on the intensity of transport activities. However, modification of infrastructure is often limited by geographical conditions or by financial reasons. The tool Villon was successful even in this area and was applied in the solution of problems following the production increase in paper production plant SCA Laakirchen, Austria, steelworks Corus Teesside in the UK and in infrastructure modification process in giant chemical plant BASF Ludwigshafen, Germany.

For a complex and detailed investigation of infrastructure of some terminals, an interaction between railway transport and other transport modes (e.g. road transport) should be considered. Villon was utilized to verify different configurations of road and railway infrastructure in Volkswagen Bratislava plant. The entire internal transport was modelled, including trucks, cars, trains and pedestrians. Austrian steel works VOEST Alpine Linz needed a simulation study of a road infrastructure reconstruction in interaction with the railway and road traffic inside the plant area.

Once built, verified and validated model of a terminal can be used repeatedly in seeking for solutions of a different type [1, 3]. For example, simulation model which was originally developed as a tool for optimal reconstruction of a terminal (strategic goal) can be also useful during the real



reconstruction process in the solution of problems caused by planned closings due to reconstruction.

8 Summary

Planning and optimisation of infrastructure and of traffic processes inside the terminals and transport nodes is unthinkable without detailed and objective evaluation of consequences which follow every decision, because extremely high money funds are at stake. The suitability of planned changes of infrastructure respectively of technological procedures cannot be evaluated objectively without a thorough study of the terminal operation after the changes are realised. Perhaps, due to extreme complexity of the terminals and their operation, the only efficient technique how to evaluate and verify all consequences of our decisions is making experiments with well fitted simulation model of the terminal.

Acknowledgements

This work has been supported by the Czech research project MSM 0021627505 "Theory of transportation systems" and by the Slovak grant project 1/1049/04 "Agent-based architecture of simulation models related to service systems".

References

- [1] Kavička, A., Klima, V.: ABASim: Agent-based architecture of simulation models, *Simulation Almanac 2005*, Czech and Slovak Simulation Society, Prague, pp.63-72, 2005
- [2] Bažant, M., Kavička, A.: Simulation model of Prague main station, *Proc. of international symposium "ŽEL 2005"*, University of Žilina, pp.18-22, 2005
- [3] Kavička, A., Klima, V., Adamko, N.: *Agent-based simulation of transportation nodes*, EDIS, University of Žilina, 2005, (in Slovak),
- [4] Adamko, N., Klima, V., Márton, P., Kavička, A.: Simulation Support of Railway Station Infrastructure Design, *Transport - Prace Naukowe*, Wydawnictwo Instytutu Technologii Eksploatacji, Radom, pp. 11-16, 2005
- [5] Márton, P., Klima, V., Kavička, A., Adamko, N.: Simulation support of railway infrastructure in logistics terminals planning, *Proc. of international symposium Transport 2004*, Univ. of Transport, Sofia, pp. 489-494, 2004
- [6] Xu, Z.L., Klima, V., Kavička, A.: The Realization of Joint Terminal Process Simulation by Computer, *Railway Transport and Economy*, China Academy of Railway Sciences, 1/2003, Beijing, pp.51-53, 2003 (in Chinese)
- [7] Kavička, A., Klima, V.: Simulations of railway station operation, *Proc. of the World congress on Railway Research*, Köln, 2001
- [8] Klima, V., Kavička, A.: Simulation support for railway infrastructure design and planning processes, *Proc. of Comrail 2000 conference*, Wessex Institute of Technology, Southampton-UK, pp.447-456, 2000
- [9] Kavička, A., Klima V., Niederkofler, A., Zaťko, M.: Simulation model of marshalling yard Linz Vbf (Austria), *Proc. Int. Workshop on Harbour, Maritime & Logistics Modelling and Simulation*, SCS, Genoa, pp.317-320, 1999



Section 12

Communications

This page intentionally left blank

Model checker for railway signalling communication protocol

J.-G. Hwang, H.-J. Jo & J. H. Lee

Train Control Research Team, Korea Railroad Research Institute (KRRRI), Korea

Abstract

As a very important part in the development of the protocol, verifications for a developed protocol specification are complementary techniques that are used to increase the level of confidence in the system functions by their specifications. Using the informal method for specifying the protocol, a little ambiguity may be contained in the protocol. This indwelling ambiguity in control systems can be the cause of accidents, especially for safety-critical systems. To clear the ambiguity contained in the designed protocol, we use the LTS (Labelled Transition System) model to design the standard protocol for railway signalling systems. Then we verify the safety and liveness properties automatically and formally through the model checking method. The modal μ -calculus, which is an expressive method of temporal logic, has been applied to the model checking method. In this paper, we verify the safety and liveness properties of the Korean standard protocol for railway signalling systems. To automatically check the safety and liveness properties of the designed protocol, the formal checker is implemented. The developed tools are implemented by the C++ language under Windows XP.

1 Introductions

A few years ago, most of the equipment consisted of vital relay-based systems ensure the safety of railway signalling systems. However, according to the computerization of railway signalling systems, lots of information is exchanged among the computerized railway signalling equipment. By the systemization of railway signalling systems, the communication link is considered more significant than before. Therefore, the communication protocol has to be clearly



defined and standardized for the systemized and intelligent railway signalling systems [1].

A new protocol for railway signalling systems has designed and standardized in our research. It is expected that the communication protocols designed by experts could have brought about some ambiguities. Provided that there were some ambiguities in the designed protocol by the experts, the ambiguities might provoke fatal flaws in the control of signalling systems or accidents. Therefore, the communication protocol for vital systems like railway signalling systems has to be correctly verified [2, 5]. The primary objectives of protocol standardization are to allow systems developed by different vendors to work together, to exchange and handle information successfully. In recent years, the application of formal methods to standardized protocol design has given rise to a new field called protocol engineering. Formal verifications are complementary techniques that are used to increase the level of confidence in the correct functioning of communication protocol by their specifications. Formal verification can give certainty about satisfaction of a required property, but this certainty only applies to the model of the specification.

For our research, we use LTS model to design the communication protocol for railway signalling, LTS model is an intermediate model for encoding the operational behaviour of processes. And then, we verify automatically and formally the safety and liveness properties through the model checking method, especially modal μ -calculus [3, 4]. This paper presents a model checking method for Korean railway signalling protocol specified in LTS and developed automatic verification tool which is able to verify formally whether properties expressed in modal logic are true on LTS specifications. The implemented formal checker using model checking method enables to verify whether deadlock and/or livelock properties are true on specifications.

This paper is organized as follows: Section 2 describes some key concepts and notations in protocol verification. Section 3 describes specifications of a Korean standard protocol for railway signalling systems. Section 4 gives the functional description of a model checker, and finally conclusions are given in Section 5.

2 Formal verification of communication protocol

2.1 Preliminaries

Communication protocols are developed through several phases, such as user requirement analysis and specification, design, implementation and conformance test phase, respectively. Most of protocol development phases are accomplished by human experts. Thus it is expected that some ambiguities or faults may be applied in protocol specification. Those included ambiguities can come into malfunction of systems or accidents. Above section mentioned, formal method are applied in protocol design phase to clearance of built-in faults or ambiguities, thus to assurance of safety and reliability of protocol for safety-critical control systems, such railway signalling systems.



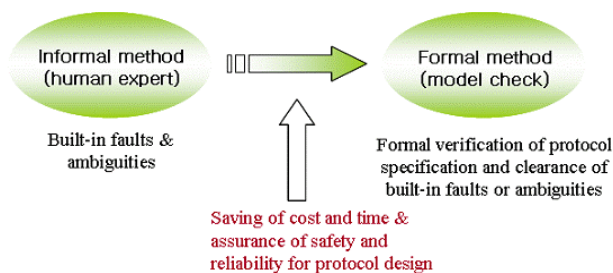


Figure 1: Introduction of formal method.

Fig. 2 shows the procedure of protocol development with formal verification. Formal verification has formal specification phase by formal description language and verification by model checking method. The inherent ambiguities or faults can be cleared and assures the safety of designed protocol through the protocol specification and this verification phase. Two shaded (yellow) colour parts in fig. 2 show the above described formal verification phase.

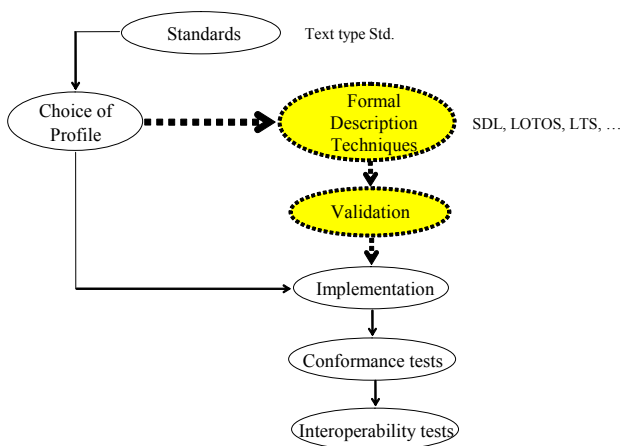


Figure 2: Protocol design with formal verification.

The protocol for railway signalling system requires more high reliability and safety than other industrial control systems. Formal verification is very useful method to verify the correctness of designed protocol. In our research, the formal verification is applied to designed standard communication protocol for Korean signalling systems. LTS model is used to design the communication protocol for railway signalling in this paper, and we verify the safety and liveness properties through the model checking method, especially modal μ -calculus. Fig. 3 shows the formal verification procedure of designed standard communication protocol for Korean railway signalling system.



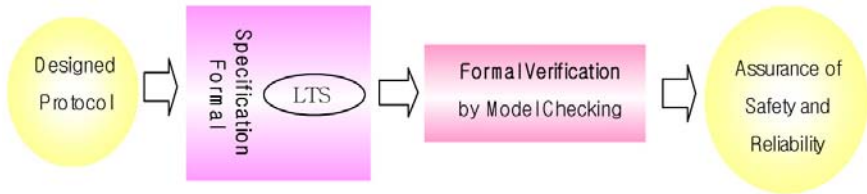


Figure 3: Formal verification procedure of railway signalling protocol.

2.2 Definition of LTS

The formalism of LTS is used for modelling the behaviour of processes, systems and components. LTS serves as a semantic model for a number of protocol specification languages, e.g. CCS (Calculus of Communication System), CSP (Communicating Sequential Processes), and LOTOS [3, 4].

Definition 1: A labelled transition system is a 4-tuple $\langle S, L, T, s_0 \rangle$ with

- S is a (countable) non-empty set of states.
- L is a (countable) set of observable actions.
- $T \subseteq S \times (L \cup \{\tau\}) \times S$ is the transition relation.
- $s_0 \in S$ is the initial state.

2.3 Model checking for verification

Model checking is a verification technique that uses formulas of a temporal logic to express properties of a system expressed in some other kind of specification language, and then matches them each other to decide whether the property holds for the system.

We use finite state LTS to specify the designed protocol specification. It has been the most common specification paradigm in recent years. Also, we choose the modal μ -calculus as property specification language. The modal μ -calculus, originally due to Kozen [4], is proposed as a highly expressive logic that can be used to specify properties of concurrent systems represented as LTS.

2.3.1 Modal μ -calculus

The modal μ -calculus can alternatively be viewed as the logic obtained by adding recursion to Hennessy-Milner logic. More generally practitioners have found Hennessy-Milner logic useful to be enabling to express temporal properties of concurrent systems. However, as logic it is not very expressive because formulas of the logic are not rich enough to express such temporal properties. Thus several operators are added in modal μ -calculus. The result is a very expressive temporal logic.



2.3.2 Safety and liveness

The protocol has two properties which it have a safety without deadlock and livelock, and a liveness with some good state and action. A safety property states that some bad feature is always precluded. Safety can either be ascribed to states, that bad states can never be reached, or to actions, that bad actions never happen. A liveness property states that some good feature is eventually fulfilled. Again it can either be ascribed to states, that a good state is eventually reached, or to actions, that a good action eventually happens.

3 Designed protocol specification

In this section, as a reference model for verification, we concentrate on a Korean railway signalling protocol between CTC communication server and EIS (electronic Interlocking System). The CTC communication server located in the centralized control centre receives the control commands from CTC main computer for the control of field signalling equipment such as signal aspects, point machines, and others. Conversely, the EIS transfers state information of the field signalling equipment CTC computer. If this link contains any faults or errors, they may lead to a severe accident because the interface link is the essential hub-link for controlling and monitoring railway signalling, so the interface link is a significant link from the point of view of safety of the railway signalling operation.

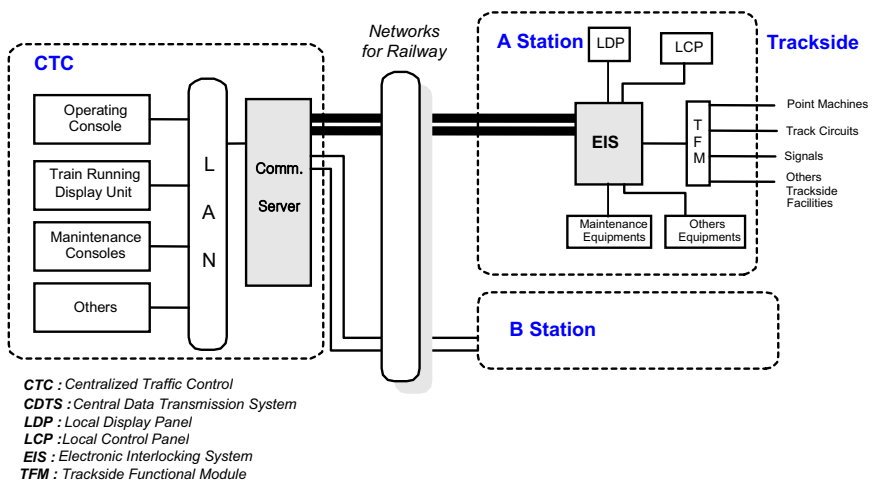


Figure 4: Configuration of railway signalling systems.

Fig. 5 specifies LTS for modelling the behaviour of railway signalling process. This LTS model has 6 states and 9 transitions. The details of standard protocol can be identified in reference [1].



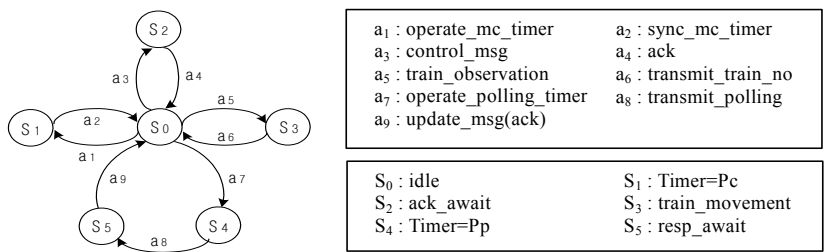


Figure 5: LTS model generated from designed protocol.

4 Model checking by formal checker

For the reference LTS shown in Figure 5, we will verify the general correctness of the LTS specification by applying the explained concepts proposed in section 2. We verify the designed protocol using the modal μ -calculus formula, which means we examine the correctness properties of the protocol. For example, the equation (1) modal μ -calculus formula has to be “true”, if the LTS of the designed protocol consists of the non-existence of deadlock and livelock. The Solve algorithm [4] as a model-checking algorithm is applied to this equation (9) formula. From this process the verifying results can be obtained.

$$\nu Z. (\mu Y. A \vee (<-> tt \wedge [-]Y)) \wedge [-]Z, A=\{S_0\}$$
 (1)

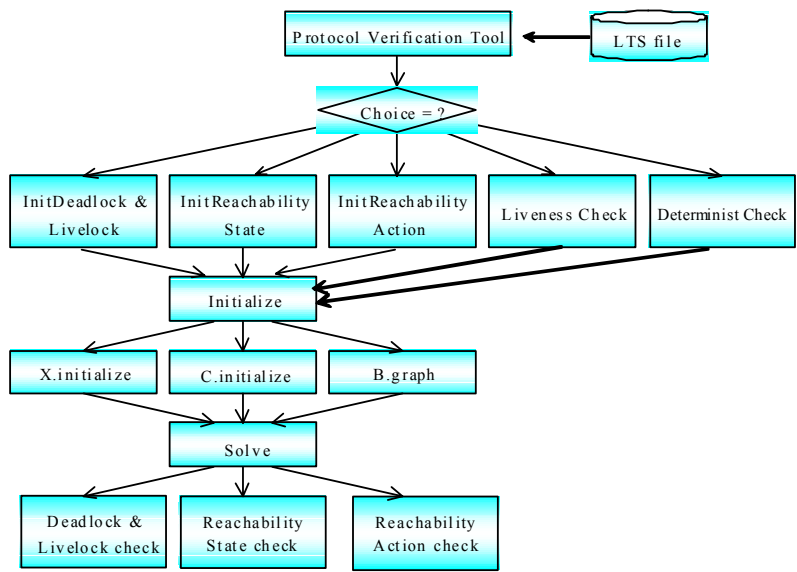


Figure 6: Operation flow of developed model checker.

We developed the formal verification tool called model checker for protocol design of railway signalling systems by above described formal specification and verification method. The verification items are deadlock, livelock, reachability, liveness and determinist properties of designed protocol. The configuration of developed model checker shows fig. 6.

The input of this model checker is the LTS model generated from designed protocol such as fig. 4. Above sections described verification algorithm, such as formal description technique, modal mu-calculus logics and Solve algorithm is implemented in model checker. The implemented formal checker is able to verify whether properties expressed in modal logic are true in specifications using modal mu-calculus. The suggested tools are implemented by C++ language under MS-Windows NT environment. In the way of fig. 7, the text-based LTS modelling file, 'lts.lts', is inputted in model checker, and clicks one of Check button what we want to verify, and then the verification results windows are popped. If the 'Solve Algorithm' is clicked, the executable process of verification by Solve algorithm becomes visible in 'Solve.txt' text file. Fig. 6 shows process of formal verification of standard protocol for Korean railway signalling systems using developed model checker.

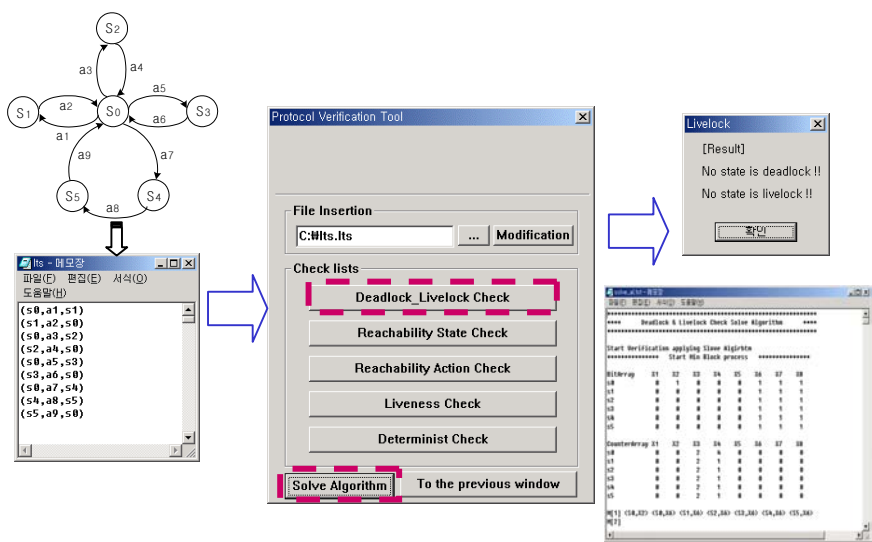


Figure 7: Deadlock and livelock checking by developed model checker.

5 Conclusion

Using the informal method in specification of the communication protocol, ambiguities are generally contained in the protocol. To clear up the ambiguity contained in the designed protocol, the protocol is specified in LTS and verified the safety and liveness properties by the model checking method. This formal verification process for designed protocol requires elaborate and difficult efforts.

In this paper, we have developed the user friendly model checker by GUI (Graphic User Interface) under MS-windows environment. It is expected to reduce the time and cost for protocol design by using this verification tool, and to increase the safety, reliability and efficiency of maintenance of the signalling systems by using the designed protocol for railway signalling in Korea.

References

- [1] J. G. Hwang and J. H. Lee, "A New Data Link Protocol for Korea Railway Signalling Systems", KIEE Int'l Trans. on EMEC, Vol.3-B, No.4, pp. 195-201, Dec. 2003.
- [2] D. Schwabe, 'Formal Techniques for the Specification and Verification of Protocol', Ph.D Thesis, Univ. of California Los Angeles, 1981.
- [3] O. Burkart and B. Steffen, Model Checking the Full Modal Mu-Calculus for Infinite Sequential Processes, LFCS Report ECS-LFCS-97-355 (1997).
- [4] Kozen, 'Results on the prepositional μ -calculus', Theoretical Computer Science, 27:333-354, December 1983.
- [5] J. H. Lee, J. G. Hwang and G. T. Park, 'Performance Evaluation and Verification of Communication Protocol for Railway Signalling Systems', Computer Standards & Interfaces in Elsevier, vol. 27, pp. 205-219, Feb. 2005.



A new ground-to-train communication system using free-space optics technology

H. Kotake, T. Matsuzawa, A. Shimura, S. Haruyama
& M. Nakagawa

*Department of Information and Computer Science, Keio University,
Japan*

Abstract

We propose a new ground-to-train communication system using free-space optics between a train and the ground. In the proposed system, a cylindrical concave lens spreads the incoming beam from transmitter (Laser Diode, LD) horizontally to form a wide fan-shaped beam. The fan-shaped beam is projected to a train and the width of the projected beam is equal to the length of a typical bullet-train car. This transmitter with cylindrical concave lens and a receiver (Avalanche Photo Diode, APD) are installed on a train and the ground, and the horizontally spread beam is received continuously by the corresponding receiver. The train can keep a communication link continuously to the ground thanks to this spread beam. We analyzed the performance of the proposed system by experiments. The experimental results show that a sufficient received Signal-to-Noise Ratio (SNR) can be obtained when a train is not moving; therefore, the signal can be received continuously even when the train is moving. Furthermore, in order to keep a continuous communication link even when the train vibrates or shifts vertically, the vertical spread angle of a laser beam is adjusted between 0.1 degrees and 0.5 degrees. These experimental results lead to the conclusion that this system is expected to be a Giga-bit class high speed communication technology between the train and the ground.

Keywords: train communication, optical wireless communication, visible light communication, horizontally spread beam.

1 Introduction

In recent years, the environment of the ubiquitous society is being developed by rapid expansion of high-speed communication infrastructure such as Asymmetric



Digital Subscriber Line (ADSL) and Fiber To The Home (FTTH). We can use many services in hotspot and internet cafe. There is a high demand for the infrastructure to provide enough service to customers in a train. Leaky Coaxial cable (LCX) and millimeter-wave have been used for wireless communication between a train and the ground. However, the data rate of LCX is only 2.56 Mbps [1]. Millimeter-wave can increase the data rate up to 1Gbps when a train is not moving, but the data rate is decreased to 6.3 Mbps when the train is moving [2]. Therefore, these systems cannot provide high-speed data transmission in the moving train.

The equipments of optical wireless communication such as LD and LED are developed rapidly, and it is possible to achieve high-speed communication with these equipments. Since frequency of lightwave is very high, optical wireless communication is suitable for high-speed communication [3, 4].

We propose a new train communication system using free-space optics technology between a train and the ground. In the proposed system, a cylindrical concave lens spreads horizontally the incoming beam from LD horizontally to form a wide fan-shaped beam. The fan-shaped beam is projected to a train and the width of the projected beam is 25 m which is equal to the length of a typical bullet-train car. Owing to this spread beam, the train can continuously keep a communication link to the ground.

In this paper, we analyze the performance of this system experimentally. We investigate the received SNR in stationary condition and received signal amplitude at moving environment. Moreover, we examine the received power with the expansion of vertical beam angle in order to have a continuous communication link. Finally, we conducted the outdoor experiment using the test train.

The rest of the paper is organized as follows. In section 2, we describe LD with variable Numerical Aperture (NA) lens and cylindrical concave lens. In section 3, we describe the proposed system model. In section 4, we evaluate the system performance experimentally. Finally, the conclusions are given in section 5.

2 Design of transmitter

2.1 LD fixed with variable NA lens

LD is a device that emits the light by induced emission and laser oscillation and outputs the coherent light. The light of LD has isolated wavelength and high directivity, and the transmission power of LD is higher than that of LED [5, 6]. Even though the light of LD has high directivity, the width of the beam projected to a train is diffused and the sufficient received optical power cannot be obtained.

To obtain sufficient power of LD at the train, a variable NA lens which can change focal distance is attached in front of the LD. Since the incident beam angle from LD can be adjusted, the transmitted power can be controlled locally. Figure 1 shows the photograph and the concept of LD fixed with variable NA lens. The vertical beam angle of this LD is set to between 0.1 and 5.72 degrees, and the horizontally spread beam angle is fixed to 4.62 degrees.



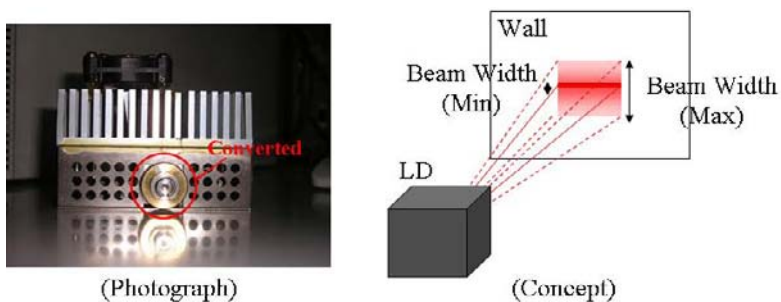


Figure 1: Photograph and Concept of LD fixed with variable lens.

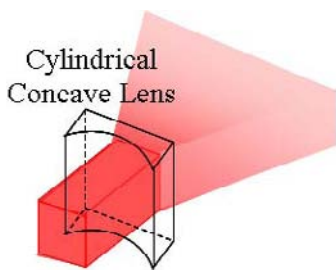


Figure 2: Effect of cylindrical concave lens.

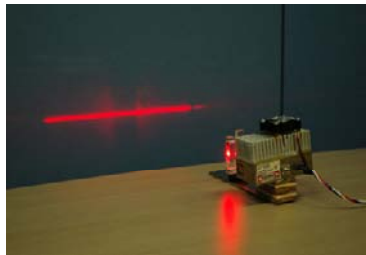


Figure 3: Photograph of laser beam which is spread by cylindrical concave lens.

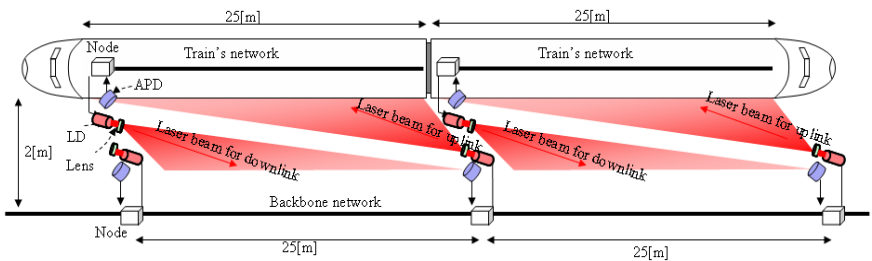


Figure 4: Proposed system model.



2.2 Cylindrical concave lens

Various types of lens are categorized by configuration. There are many types of lens such as spherical lens, paraboloidal lens, and cylindrical lens. Among various types, cylindrical lens have a shape of cylinder, and can bring about an effect of lens in one direction. Since horizontally spread beam is projected to a train in the proposed system, we select cylindrical concave lens which spreads horizontally the incoming beam from LD to form a wide fan-shaped beam. Figure 2 shows the effect of cylindrical concave lens, and Figure 3 shows the photograph of a laser beam which is spread by the cylindrical concave lens. In this paper, we select three beam angles spread horizontally by cylindrical concave lens such as 20, 30, and 40 degrees.

3 Description of proposed system

3.1 System model

The proposed system model is illustrated in Figure 4. The width of the projected beam is 25 m which is equal to the length of a typical bullet-train car. Since the horizontal distance between a bullet-train and LCX is 1.9 m, if it is possible to communicate at this distance, LD and APD can be installed instead of LCX. Because of this, the horizontal distance between train and ground is 2 m (the height of LD and APD is the same). The down-link communication from the ground to the train and the up-link communication from the train to the ground use the same transmitter (LD) and receiver (APD) so that the speed of down-link and up-link are designed to be the same.

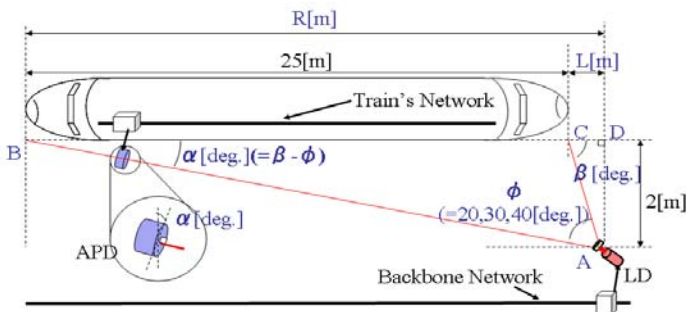


Figure 5: Proposed layout model.

3.2 Layout model

The proposed layout model is shown in Figure 5. Figure 5 shows only down-link, but up-link is also designed in a similar manner. We calculate various parameters to set up the layout. The calculated values of the various parameters with the width of the projected beam of 25 m are listed in Table 1, where ϕ is

horizontally spread beam angle and is set to 20, 30, 40 degrees, α is the angle which received at the longest point B and β is the angle which received at the shortest point C. In other words, α is the angle at which APD is tipped. L is the distance between point C and D, and R is the distance between point B and D.

Table 1: Parameters of layout model.

φ (degrees)	40	30	20
α (degrees)	4.3	4.1	3.9
β (degrees)	44.3	34.1	23.9
L (m)	2.05	2.95	4.51
R (m)	27.05	27.95	29.51
AB (m)	27.12	28.02	29.58
AC (m)	2.86	3.56	4.93

Table 2: Experimental parameter.

Bit rate	100 (Mbps)
Transmission power of LD	100 (mW)
Wavelength of LD	670 (nm)
Frequency of LD	400 (MHz)
Horizontally spread beam angle (with lens)	20/30/40 (degrees)
Vertical beam angle	0.1-5.72 (degrees)
Receiver	APD
Diameter of receiver	0.5 (mm)
Material of cylindrical concave lens	Synthetic fused silica
Condition of room	Dark room
Received signal amplitude of ambient noise	4 (mV)

4 Experimental results

In this section, we evaluated the system performance experimentally. In particular, we show the received SNR in a stationary condition, received signal amplitude in a moving condition, the received power with the expansion of vertical beam angle. Table 2 shows experimental setup.

4.1 Received SNR in a stationary condition

Measurement environment and result of the received SNR in stationary condition are shown in Figure 6. With φ as a parameter, LD is installed in order to make the width of the projected beam to 25 m, and the tipped angle of APD is 4 degrees. We measured the received SNR between 1 m point and 25 meter point at 1 m intervals using a miniature train. It can be seen from this result that even though the received SNR decreases gradually as the distance becomes longer, sufficient received SNR level can be obtained at 25 m point irrespective of φ .



Furthermore, we can find that optimal value of ϕ is 20 degrees. Therefore, we can say that a sufficient SNR can be obtained for the 25 m train length.

4.2 Received signal power in a moving condition

Measurement environment and result of the received signal power at moving situation are shown in Figure 7 with $\phi = 20$ and the tipped angle of APD of 4. We measured the received signal amplitude in a moving condition between 1 meter point and 25 m point using a miniature train. It can be seen from this result that the received signal amplitude declines gradually when the train is moving. However, it is confirmed that the signal amplitude can be received continuously for the communication area.

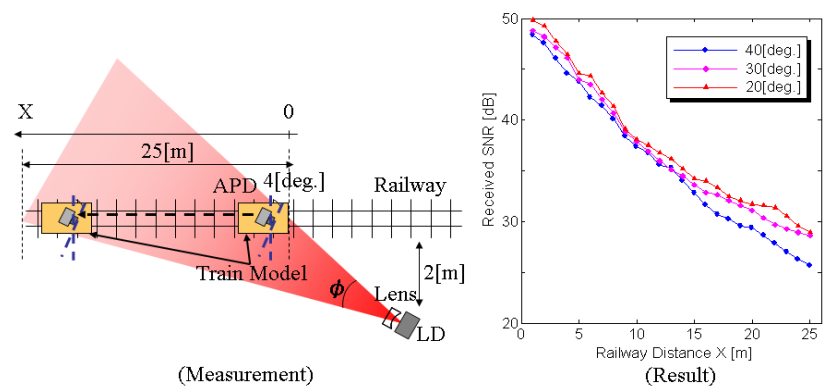


Figure 6: Measurement and result of the received SNR in a stationary condition.

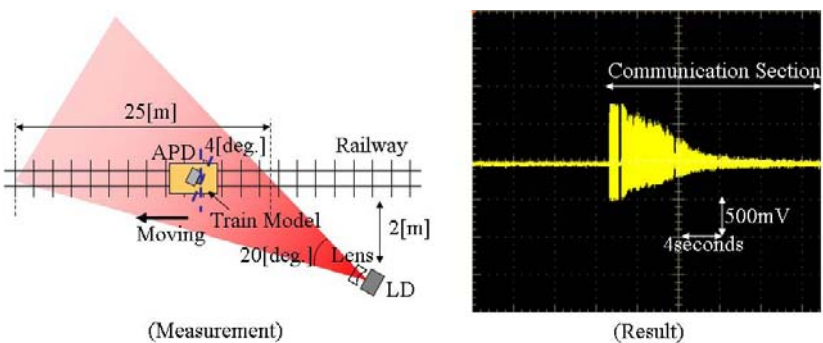


Figure 7: Measurement and result of the received signal amplitude in a moving condition.

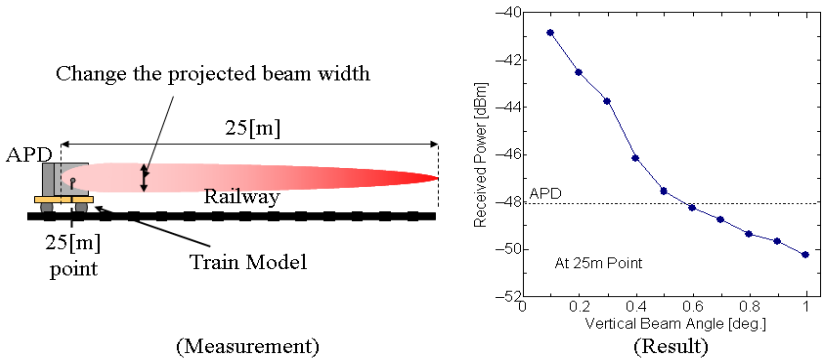


Figure 8: Measurement and result of the received power at 25 m point with the expansion of vertical beam angle.

4.3 Received power with the expansion of vertical beam angle

When a train vibrates and shifts vertically, the sensor of APD sometimes jolt out of alignment from the projected beam. We measured received power at 25 m point with expanding the vertical beam angle of LD. Measurement environment and result of the received power at 25 m point with the expansion of vertical beam angle are shown in Figure 8 with $\phi = 20$ and the tipped angle of APD of 4. A dotted line in Figure 8 shows the minimum sensitivity of APD assuming 1 GHz optical signal. It can be seen from this result that at the vertical beam angle of 0.5 degrees the sufficient optical signal can be received. Therefore, we can say that by adjusting the vertical beam angle between 0.1 and 0.5 degrees, the vertical vibration and shift of a train can be absorbed.

4.4 Outdoor experiment using test train

We conducted the outdoor experiment using test train. We measured the received signal amplitude in the down-link and up-link when the train is moving at a speed of 15-20 km/h. Measurement environments of the outdoor experiment using test train are shown in Figure 9. Photograph of the outdoor experiment is shown in Figure 10. Table 3 shows experimental setup. In this experiment, a lens which works as optical concentrator and an interference filter are attached in front of APD. The length of platform is 20 m and that of test train is 19.6 m. The width of the projected beam is 25 m, the horizontal distance between LD and APD is 2 m. $\phi = 20$ degrees, and the tipped angles of APD is 14 degrees because Full Width at Half Maximum (FWHM) of lens is 17.3 degrees. Furthermore, the vertical beam angle is 0.3 degrees. Figure 11 shows the result of the outdoor experiment. It can be seen from these results that the received signal amplitude declines suddenly at the certain time when the train is moving. We think that these results from effect of FWHM of lens on sensor of APD and interference



filter. However, it can be seen that the signal amplitude can be received continuously for 25 m. Therefore, from these results, we can say that the proposed system will be a promising candidate for train communication.

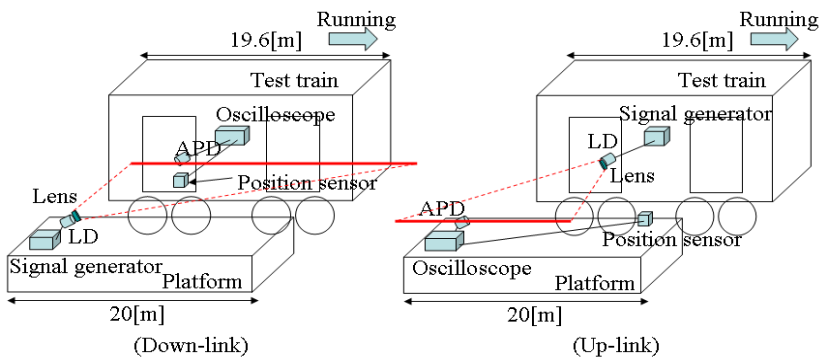


Figure 9: Measurements of the outdoor experiment using test train.



Figure 10: Photograph of the outdoor experiment.

Table 3: Experimental parameter of the outdoor experiment.

Bit Rate	100 (Mbps)
Transmitted power of LD	100 (mW)
Wavelength of LD	670 (nm)
Frequency of LD	400 (MHz)
Horizontally spread beam angle (with lens)	20 (degrees)
Vertical beam angle	0.3 (degrees)
Receiver	APD with lens
Angle of receiver	14 (degrees)
Diameter of receiver	0.5 (mm)
Material of cylindrical concave lens	Synthetic fused silica
Weather	Fine day
Received signal amplitude of ambient noise	20 (mV)

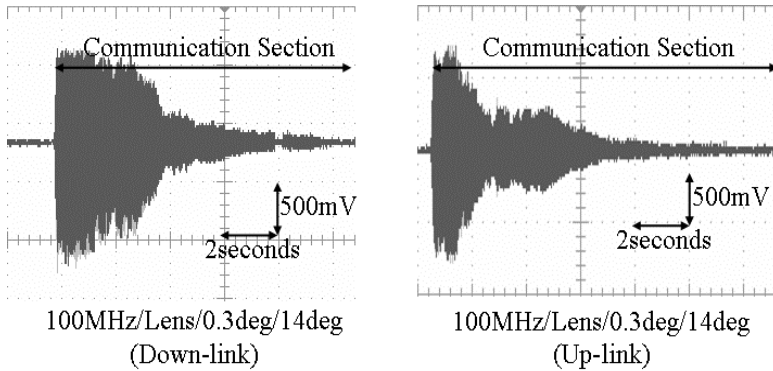


Figure 11: Results of the outdoor experiment.

5 Conclusions

In this paper, we proposed a new ground-to-train communication system using free-space optics technology in order to increase the transmission rate between a train and the ground. We investigated the system performance experimentally. It was shown from the experimental results that a sufficient SNR at stasis could be obtained for 25 m train length, and it was found that the optimal horizontal beam angle is 20 degrees. Moreover, the signal amplitude was more sufficient than the signal amplitude of ambient noise and a train can keep a communication link continuously for a 25 m train length. Furthermore, it was shown that by adjusting the vertical beam angle between 0.1 and 0.5 degrees, the effect of the vertical vibration and a shift of a train can be absorbed. Finally, we conducted the outdoor experiment using a test train. We believe that the proposed system is a promising candidate for train communication.

References

- [1] Jun, H.W., & Mei, K.K., Theory and analysis of leaky coaxial cables with periodic slots. *Antennas and Propagation, IEEE Transactions on*, Vol. 49, No.12, pp. 1723-1732, 2001.
- [2] Nikura, H., Akiyama, Y., & Adachi, D., Train radio system with millimeter wave used for the Yamanashi maglev test line. *RTRI Report*, Vol. 15, No. 11, 2001.
- [3] Komine, T. & Nakagawa, M., Integrated system of white LED visible-light communication and power-line communication. *Consumer Electronics, IEEE Transactions on*, Vol. 49, No.1, pp. 71-79, 2003.
- [4] Tanaka, Y., Komine, T., Haruyama, S. & Nakagawa, M., Indoor visible light communication utilizing plural white LEDs as lighting. *Personal, Indoor and Mobile Radio Communications*, Vol. 2, pp. F-81-F-85, 2001.



- [5] Kahn, J.M. & Barry, J.R., Wireless infrared communications. *Proc, IEEE*, Vol. 85, No. 2, pp. 265-298, 1997.
- [6] Barry, J.R., *Wireless infrared communications*, Kluwer Academic Publishers, Boston, MA, 1994.



Communications security concerns in communications based train control

M. Hartong¹, R. Goel² & D. Wijesekera¹

¹*George Mason University, USA*

²*Howard University, USA*

Abstract

Since the late 1980s, Communication Based Train Control (CBTC) Systems for freight and passenger rail have been under development in the United States. These systems have been advertised as offering significant enhancements in safety by ensuring positive train separation, enforcing speed restrictions, and improving roadway worker protection. In order to maximize the effect of these safety enhancements, it is necessary for CBTC systems to address security issues common to wireless computer communication systems. This paper introduces the role that CBTC systems play in railroad methods of operations, as well as the vulnerabilities of communications systems being manifested in CBTC. It provides a classification of attacks against CBTC systems, and identifies the security controls to mitigate these attacks.

The level of risks associated with these security issues have increased from the first CBTC system introduction, primary because of increases in the means of exploiting the associated vulnerabilities. Exploitation that could compromise the system safety capabilities can take the form of any number of different types of attacks (e.g. jamming, etc). Failure of the CBTC system designer to adequately address these attacks could allow a malicious party to exploit CBTC vulnerabilities, effectively neutralizing the safety advantages of a CBTC system. Recent non-CBTC train-to-train collisions causing release of toxic inhalants and resulting deaths illustrate that a lack of CBTC system safety capabilities could have catastrophic results. Fortunately, these attacks can be mitigated using various security controls. Understanding the attacks and the respective mitigating security mechanisms is therefore key to effectively implementing CBTC safety advantages.

Keywords: security, communications based train control, wireless, communications systems, positive train control.



1 Introduction

Today's railroads are a critical component in the US transportation and distribution system. US freight railroads have grown in 2003 to a \$38 billion dollar industry with 549 freight railroads and 141,000 miles of track. In 2001, the last year for which data has been calculated, that equated to 25% of all intercity freight tonnage carried in the US, and 41% of all ton-miles. This is approximately 12% of all freight revenue in the US in ton-miles, making the rail industry the premier low cost competitive service to other forms of transportation and distribution [1]. Today modern diesel electric locomotives in the US routinely pull over 100 freight cars weighing 286,000 pounds at speeds up to 60 miles per hour. These sizes, speeds, and cargo capacity can result in significantly adverse consequences if the existing methods of train operation fail.

Communication Based Train Control (CBTC) Systems offer protections against the failure of existing methods of train operation. In order to realize these protections, however, it is necessary for CBTC systems to address security issues common to wireless computer communication systems. This paper introduces the role that CBTC systems play in railroad methods of operations, as well as the vulnerabilities of communications systems being manifested in CBTC. It provides a classification of attacks against CBTC systems, and identifies the security controls to mitigate these attacks.

2 Methods of operations and traditional train control

In order to control the movement of trains, various methods of operations began to be formalized starting in the early 1820s when multiple trains began to share the same set of tracks. These methods of operations were designed to improve the operational efficiency and safety of the railroad through the reduction of collisions, derailments, and the associated deaths. Today methods of operations for the control of trains can be classified in to four basic categories: verbal authority, mandatory directives, signal indications, and signal indications supplemented by cab signals, automatic train control, or automatic train stop systems. CBTC systems support these modes of operation.

2.1 Verbal authority and mandatory directives

With verbal authority and mandatory directives, the aspect of wayside signals does not control train operations. Instead, train operations are controlled by orders from the Train Dispatcher, who takes responsibility for knowing what trains are located where, and ensures that no two trains are issued authorization to occupy the same location of track at the same time. The Dispatcher usually issues orders, mandatory directives, speed restrictions, as well as the location of any wayside work crews via two-way radio to the locomotive crew. The train crew then is responsible for ensuring that they obey these orders, speed restrictions, and advisories.



This is the traditional means of controlling operations in the United States, and roughly 40% of all tracks in the United States are controlled in this manner. Verbal Authority and Mandatory Directives operations are generally broken either one of two main types- Track Warrant Control (TWC) and Direct Traffic Control (DTC). TWC and DTC do not require wayside signals. They can, however be used to supplement Automatic Block Signalling (ABS) to increase flexibility and traffic capacity.

When used as a supplemental mode of operation, DTC/TWC serves primarily as a protective overlay to the movement authority and do not convey the authority to occupy the main track. That authority remains with the signal system, but the train crew requires a DTC/TWC authorization in addition to the signal to enter the main track. In TWC the verbal instructions are given for the crew to proceed between stations or mileposts (a segment of track known as the authority limit). DTC is similar to TWC, but because the railroad is divided into pre-defined “blocks,” it is simpler in execution. Movement authorities can only be specified in terms of the pre-defined blocks. The decision to use TWC or DTC is made by individual railroads based on what is most efficient for their operations.

2.2 Signal Indications

Train operations under signal indications makes up the remainder of the train control operations in the US. Track circuit based signal systems were first installed in the US in 1872, and in 1927 were centrally controlled in the first “Centralized Traffic Control (CTC)” system. CTC is not a separate control system—it uses block signal system and interlocking to control train movements (although radio communications between the dispatcher and train crews are available).

CTC, sometimes called Traffic Control System (TCS) has remained basically unchanged since the 1930s. In CTC authority for train movements are provided by signal indications. The train dispatcher at the control centre determines train routes and priorities, and then remotely operates switches and signals to direct the movement of trains. The CTC system is designed so that the dispatcher cannot grant conflicting authorities.

Some CTC systems have been enhanced to provide direct indications of wayside signals aspects to the locomotive engineer inside the locomotive cab. These “cab signal” systems provide on-board display of trackside signal indications through the transmission of signal aspect information in coded pulses along the track. The engineer controls the speed of the train with the signal information, and obtains authority to enter sections of track. Further refinements called “automatic train stop” or “automatic train control” systems automatically cause the train to stop or reduce speed where an engineer fails to respond appropriately to a trackside signal.

2.3 Limitations of Current Train Control Technologies

Cab signals simply relay the external signal indications to a visual display inside of the cab of the locomotive, making it easier for the crew to note the signal



aspect and the associated order it conveys. Unless operated with ATS or ATC, the cab signal systems do not provide speed or authority enforcement. Consequently, no mechanism would exist to detect and prevent crew non-compliance with dispatcher orders and railroad-operating procedures.

ATS provides enforcement for signal indications. This can be done with or without a cab signals system in place. ATS however, does not provide speed enforcement. It only enforces the indication provided by the wayside signal in the event that the train crew fails to react. ATC, on the other hand, provides both signal indication enforcement as well as speed enforcement.

In general, ATS and ATC systems have several significant technical limitations. First, the location of trains can only be determined to the resolution of the track circuits. The track circuit's length can be made shorter, but adding additional track circuits requires additional wayside hardware. This imposes additional costs, causing a practical (and economical) limit to the number of track circuits that a railroad can install. Second, the information that can be provided to a train through a rail based is limited to a small number of wayside signal aspects or speed data.

In addition, the underlying signal systems to provide the required indications for cab, ATS, or ATC to operate are capital intensive. In 2003, the US Class 1 railroads alone spent over \$490 million in operations, administration, and maintenance of all types of communications and signalling systems with another \$153 million in depreciation of the existing plant [2] on approximately 65,000 miles of track. Consequently the deployment of these technologies is limited to those areas where rail throughput needs to be maximized. Less than 5% of route-miles in the US [2] have systems in place where signal indications are shown in the locomotive cab or there is on-board enforcement of the signal indications, or both.

3 Communications Based Train Control (CBTC) and methods of operation

Positive Train Control (PTC) is a name applied to select CBTC systems capable of supporting, at a minimum, the following three functions [3]

- (a) Prevention of train-to-train collisions (positive train separation).
- (b) Enforcement of speed restrictions, including civil engineering restrictions and temporary slow orders.
- (c) Protection for roadway workers and their equipment operating under specific authorities.

The inability of cab signals, ATS, and ATC to effectively incorporate collision and accident avoidance measures with the current methods of operations has been the primary motivation for the US National Transportation Safety Board (NTSB) call for PTC [4]. These CBTC systems can overcome the fundamental limitations of conventional ATS and ATC Systems.

In addition to classification by functionality, PTC systems are also classified by the extent that they used to augment the existing method of railroad operations. "Full" PTC Systems do not simply augment the existing mode of



operation with their functionality, but change or replace it. “Overlay” PTC systems provide their functionality while maintaining the existing method of operation.

3.1 Generic architecture

The generic PTC functional architecture consists of three major functional subsystems. These are:

- (a) The wayside units, consisting of elements such as highway grade crossing signals, switches and interlocks or maintenance of way workers,
- (b) The mobile units, which are the locomotives or other on rail equipment with their onboard computer and location systems, and
- (c) The central office dispatch/control unit.

All of the subsystems are interconnected by communications links.

Each major functional subsystem is a set of physical components implemented using various databases, data communications systems, and information processing equipment. The physical components that make up each subsystem depend upon the functional capabilities of each subsystem. Varying degrees of subsystem functional capability can result in significantly different hardware and software equipment & configurations.

4 CBTC system vulnerabilities

Recent research has examined security and possible problems in the rail infrastructure [5] and surveyed systems in use [6]. Completion of recent regulatory initiatives [7], coupled with accelerated industry efforts in the deployment of CBTC systems [8], have increased the level of risk that the public may potentially be exposed to as a result of the greater use of wireless technology. The most significant source of risk in wireless networks is that the technology’s underlying communications medium, the airwave, is open to intruders.

Changes in malicious hacker activity have shifted from conventional fixed wired systems to wireless networks. These networks have included not only traditional telecommunications systems, but also industrial control systems. Studies by the National Research Council and the National Security Telecommunications Advisory Committee [9] show that hacker activity includes the ability to break into wireless networks resulting in the degradation or disruption of system availability. A recent General Accountability Office study [10] has indicated that successful attacks against control systems have occurred. While these studies were unable to reach a conclusion about the degree of threat or risk, they uniformly emphasize the ability of hackers to cause serious damage.

The resources available to potential intruders are significant [11]. Intelligence is already widely available on the Internet that enables intruders to penetrate any sort of traditional computer network and wireless systems. Detailed vulnerability information is publicly discussed on newsgroups. Tutorials are



available that describe how to write automated programs that exploit wireless systems vulnerabilities. Large numbers of automated software tools have been written that enable anyone to launch these types of attacks. Publicly available Web sites whose sole purpose is to distribute this data have been established, often ensuring wide spread distribution of the information before public access can be terminated.

4.1 Attacks

The Information Assurance Technical Framework Forum (IATFF), an organization sponsored by the National Security Agency (NSA) to support technical interchanges among U.S. industry, U.S. academic institutions, and U.S. government agencies on the topic of information assurance, has defined five general classes of information assurance attacks- passive, active, close-in, insider, and distribution [12]

4.1.1 Passive attacks

The danger of a passive attack is a result of the surreptitious way information is gathered. It is the easiest type of attack to execute, and the hardest to defend against. Since the attacker is not actively transmitting or disturbing the transmitted signal of the signal owner, the signal owner (defender) has no means of knowing that their transmission has been intercepted. This kind of attack is particularly easy for two reasons: 1) frequently confidentiality features of wireless technology are not even enabled, and 2) because of the numerous vulnerabilities in the wireless technology security, determined adversaries can compromise the system.

4.1.2 Active attacks

Active attacks that can be launched against a wireless network come from a broad continuum. In its simplest form, active attacks use some mechanism disabling the entire communications channel between the sender and the receiver. With the original sender and receiver unable to recognize transmissions between each other, they cannot exchange information, and are unable to communicate. No detailed knowledge of the message parameters between sender and receiver is required, only a device capable of blocking communications operating over the entire channel.

More sophisticated forms of active attack are the “Denial of Service (DOS)” or the more advanced “Distributed Denial of Service (DDOS). The DOS and the DDOS differ primarily in the location of the origin of the attacks. The DOS originates from only one location, the DDOS from multiple locations. The specific mechanisms of a DOS and DDOS are very communications protocol and product implementation dependent, since these attacks exploit weaknesses in both the communications protocol and the products implementation of the protocol.

Other active attacks are based on exploitation attempts associated with the sender (identity theft, where an unauthorized user adopts the identity of a valid sender), weakness associated with the receiver (malicious association, where



unsuspecting sender is tricked into believing that a communications session has been established with a valid receiver, or weaknesses associated with the communications path (man in the middle, where the attacker emulates the authorized receiver for the sender- the malicious assertion, and emulates the authorized transmitter for the authorized sender- identity theft.). These attacks are primarily geared at disrupting integrity in the form of user authentication (assurance the parties who they say they are), data origin authentication (assurance the data came from where it said it did), and data integrity (assurance that the data has not been changed).

4.1.3 Close in, insider and distribution attacks

These three categories describe the nature of system access, as opposed to the passive or active nature of the attack. Close-in, insider, and distribution attacks make use of some form of either an active or passive attack whose effectiveness is enhanced by the degree of the attackers' access to the system. Insider and distribution attackers usually will utilize their specialized knowledge or access to carry out some form of a passive or active attack.

4.2 Attack mitigation

The basic security mitigations for information and information processing systems attacks in the United States have been codified in law [13]. Specifically these are confidentiality, integrity, and availability. Confidentiality is concerned with ensuring that the data and system are not disclosed to unauthorized individuals, processes, or systems. Integrity ensures that data is preserved in regard to its meaning, completeness, consistency, intended use, and correlation to its representation. Availability assures that there is timely and uninterrupted access to the information and the system.

Closely related to these three are authenticity, accountability, and identification. Authenticity is the ability to verify that a user or process that is attempting to access information or a service is who they claim to be. Accountability enables events to be recreated and traced to entities responsible for their actions. Authenticity and accountability require the ability to uniquely identify a particular entity or process, as well as the authorizations (privileges) that are assigned to that entity. Identification is the specification of a unique identifier to each user or process.

4.2.1 Countermeasures for passive attacks

In general, the preferred mitigation methods for passive attacks are access control and confidentiality. Access control mechanisms are used to prevent unauthorized users accessing services and resources for which they have not been granted permission and privileges as specified by a security policy. Confidentiality can prevent the gain of information about from the content of the messages exchanged.

4.2.2 Countermeasures for active attacks

In general, the preferred mitigation methods against active attack include access control, availability, accountability, authentication, and integrity. The access



control and availability countermeasures must maintain or improve data availability. The system must be able to ensure the availability of both data and services to all components in the system. In the event that a PTC platform cannot handle its computational and communication load, it must provide graceful degradation of services and notify the operator that it can no longer provide the level and quality of service expected to prevent an unintentional denial of service.

Ensuring integrity (and confidentiality) places restraints on availability and has performance costs. Encrypting agents and messages in transit may impose unacceptable delays in environments where near real-time response is required.

The use of Cyclic Redundancy Codes (CRCs) is sometimes claimed as a means of providing data integrity. A CRC does not provide protection malicious errors. This is the result of the CRC many to one relationship between its input and output. It is possible for multiple inputs to check sum to a single CRC value. As a result a data substitution can be made, with a correct CRC, and remain undetected. A cryptographic hash functions, where there is a unique one to one relationship between input and output, and where only one data input can check sum to one hash value. Any change in the input results in a change in the hash value, which is detected at the receiver when the hash calculation is carried out and the received hash value does not correspond to the calculated hash value. [14]

Authentication mechanisms provide accountability for user actions. User authentication and data origin authentication differ in that user authentication involves corroboration of the identity of the originator in real time, while data origin authentication involves corroboration of the source of the data (and provides no timeliness guarantees). User authentication methods range from so called time invariant “weak” authentication methods such as simple passwords to time variant “strong” cryptographically based authentication methods. In non-hostile environments no or weak user authentication may be acceptable, while in hostile environments strong user authentication is essential to provide authenticity. Data origin authentication provides assurances regarding both integrity and authentication. They rely on the use of digital signatures and can be either symmetrical or asymmetrical digital signature methods.

5 Summary and future work

This paper has introduced the basic architectures of Positive Train Control. It has highlighted key security issues associated with wireless PTC, and identified the main security requirements that PTC systems must possess. It has also introduced the network security requirements for PTC systems and highlighted basic security interoperability issues.

There are significant areas remaining for future study. Additional work needs to be undertaken to detail the various security architecture requirements of PTC with possible alternative techniques and technologies. The entire area of security network management for PTC systems requires further study, in terms of both



policy and technology. Potential interoperability policy and technical issues represents an additional field for further study.

Basic PTC systems, although they are economically unviable in terms of their safety case alone [5, 15], may, when combined with other advanced technologies, potentially offer significant societal benefits [16]. This success, however, will depend, upon the ability to rely on the transmitted information, which will be a function of the security that can be provided.

References

- [1] Association of American Railroads, Policy & Economics Department, US Freight Railroad Statistics, 28 October 2004
- [2] US Surface Transportation Board, Office of Economics, Environmental Analysis and Administration "Statistics of Class I Freight Railroads in the United States 2003"
- [3] Federal Railroad Administration Railroad Communications and Train Control Report to Congress, July 1994
- [4] US National Transportation Safety Board, "NTSB Most Wanted Transportation Safety Improvements 2004-2005", November 2004
- [5] Carlson, Frincke, Laude, "Railway Security Issues: A Survey of Developing Railway Technology", Proceedings of the International Conference on Computer, Communications, & Control Technology, International Institute of Informatics and Systematics, 2003.
- [6] Craven, "A Brief Look at Railroad Communication Vulnerabilities", Proceedings 2004 IEEE Intelligent Transportation Systems Conference Washington, D.C
- [7] National Archives and Records Administration, "49 CFR Parts 209, 234, and 236 Standards for the Development and Use of Processor Based Signal and Train Control Systems; Final Rule", Federal Register, 7 March 2005
- [8] National Transportation Safety Board Positive Train Control Systems Symposium March 2-3, 2005 Ashburn, Virginia
http://www.ntsb.gov/events/symp_ptc/symp_ptc.htm
- [9] The President's National Security Telecommunications Advisory Committee Wireless Task Force Report "Wireless Security" January 2003
- [10] United States General Accountability Office, GAO Testimony Before the Subcommittee on Technology Information Policy, Intergovernmental Relations and the Census, House Committee on Government Reform "Critical Infrastructure Protection Challenges and Efforts to Secure Control Systems", Tuesday, March 30, 2004
- [11] Investigative Research for Infrastructure Assurance Group, "Diversification Of Cyber Threats", Institute For Security Technology Studies At Dartmouth College, May 2002
- [12] US National Security Agency Information Assurance Solutions, Fort Meade, MD, "Information Assurance Technical Framework (IATF), Release 3.1, Sept 2002, ,



- [13] Federal Information Security Management Act of 2002 (Public Law 107-347) December 2002.
- [14] Menezes, van Oorschot, Vanstone, "Handbook of Applied Cryptography" 5ed, CRC Press, Aug 2001
- [15] Federal Railroad Administration, "Report of the Railroad Safety Advisory Committee to the Federal Railroad Administrator, Implementation of Positive Train Control Systems" August 1999
- [16] Federal Railroad Administration, "Benefits and Costs of Positive Train Control, Report in Response to Request of Appropriations Committees", August 2004



Ethernet-based network with high reliability for railway signaling systems

H.-J. Jo, J.-G. Hwang & Y.-K. Yoon

Train Control Research Team, Korea Railroad Research Institute (KRRRI), Korea

Abstract

Railway signaling systems are computerized equipment vital for the guarantee of the safe running of trains. Because of the computerization of railway signaling systems, the importance of a networking interface between these pieces of equipment is increasing. Thus, it is important to have a reliable and safe communication link between signaling equipment. Recently, network technologies have been applied in a real-time industrial control system, and also there are several studies applying computer network technology in vital control systems such as railway signaling systems. To deploy these computer network related studies in railway signaling systems, implementation costs, transmitted reliability, safety assurance technique and compatibility etc, are considered. The computer network protocol for railway signaling systems has an important advantage over the widely used Ethernet in terms of deterministic characteristics. In this paper, we propose computer network technology, especially the Ethernet, for railway signaling systems. The computer network characteristics of real-time industrial control systems and vital railway signaling systems are presented in this paper. This algorithm presented is based on switched Ethernet technology with a redundancy scheme. Also we demonstrate the experimental results of the proposed network algorithm.

1 Introduction

Railway signaling systems are computerized vital systems for the guarantee of the running trains safely which take charge of controlling train speed and direction, especially preventing train collisions. Every signaling system plays



each role, and is linked with other signaling systems, which collectively form one big signaling system.

The links between controlling equipment for railway signaling have some problems with maintenance and repairs and on adding new equipment because the way of linking is point-to-point communication. Recently, railway advanced countries are trying to study and develop ways of linking signaling systems by networking them, and some of the countries have already used the network method [1, 2]. The advent background of studying railway signaling system networks, using science technology, is an increasing need of networking for transmitting data between railway signaling systems.

Industrial network technologies were earnestly developed from the middle of the 1980s, and network technologies for Ethernet-based industrial control systems, such as RETHER, 3ComPACE, HP AnyVG LAN, IEEE 802.1p which are MAC class mechanism for supporting the real-time traffic, are studied. To apply industrial networks to railway signaling systems, choosing appropriate protocols should include considering various elements. On a broad view, cost, data transmission rate, extensibility, supporting media, the ways of media connection, and other elements should be considered. Considering these elements, the network of railway signaling systems is consequently determined to be developed by Ethernet-based protocol.

Ethernet appears as the next generation industrial network because of its open structure and low-cost. Recently, the switched Ethernet shows a very promising prospect for industrial networking because the switching technology can eliminate frame collisions. Ethernet applications to industrial controlling systems was impossible because the real-time requirements were not satisfied owing to the random characteristics of the Ethernet protocol, but the recently advanced switched Ethernet enables the Ethernet technology to be applied to industrial controlling systems [4]. By the theoretical explanations of the switched Ethernet, we identify the possibility of applying it to railway signaling systems.

In this paper, we analyze the performing efficiency through an Ethernet-based network algorithm which has high reliability, including the algorithms for detecting deficiencies and its recovery when network defects occurring. If the proposed Ethernet-based network algorithm is applied to interfaces between railway signaling systems, greater development of the extension and the maintenance and repairs of the signaling control systems can be expected, and so does the consistent operation of the signaling control systems, accompanied by a guarantee of the communication system's reliability.

2 Mathematical analysis of Ethernet for railway signaling systems

If these Ethernets are directly applied to the railway signaling systems, and transmitted data do not have regular transmission delays, then the Ethernets problematically have random transmission delays namely, uncertainty of transmission delays. And, the methods to reduce the number of collisions in Ethernet systems also have problems, because people should modify data link



classes or TCP/IP classes directly. In this paper, one problem of the Ethernet applications is solved by bringing the switching technologies as one way to solve the uncertainty of transmission delay in the Ethernet. The switching technologies were mainly developed and applied to the office network, but their trials for applications to industrial networks are increasing, as their generalization and cost decrease.

2.1 Ethernet and switched Ethernet

IEEE 802.3, often referred to as Ethernet, was developed for data communications among computers in the early 1970s by the IEEE, and is the basis of a physical layer and a data link layer of the office communication. The CSMA/CD (Carrier Sensing & Multiple Access/Collision Detection) is the media access control method of IEEE 802.3. In the Ethernet, the more there is of traffic by the CSMA/CD algorithm, the more collisions occur, so the transmission delays are increasing by waiting for the backoff time which are as long as the times of collisions. And also, the delaying time cannot be predicted because of the delaying by the random time following the BEB algorithm [4].

The switched Ethernet differs from the conventional one. In the switched Ethernet, collisions between stations are prevented because the dedicated virtual circuits between the stations, which communicate through switches, are installed. Collisions do not occur in those cases of which many stations transmit simultaneously because the switches send frames only to the determined destination stations when the source stations transmit frames. Also, collisions do not occur before the frames are not received because the full duplex method is used in the switched Ethernet, which performs by sending and receiving by each line.

A typical method of switching technology is the store and forward method. The switch receives a frame through a transmission line from a source station, and then checks if the reception line of a destination station is idle. If the reception line is idle, the switch transmits the frame. Otherwise, the switch stores the frame into its buffer and waits until the reception line becomes idle. In addition, if several frames with the same destination address are received at the switch simultaneously, the switch stores frames to the buffer and then sends frames to the destination one by one.

2.2 Theoretical analysis of maximum transmission delay for Ethernet and switched Ethernet

The performance analysis of Ethernet has not yet been exactly achieved because of the probability characteristics of the BEB algorithm. In this paper, we theoretically analyze the maximum transmission delay of CSMA/CD, as one of the analysis methods by using the timing analysis of data link classes [3]. The maximum transmission delay of the existing Ethernet occurs in the following three kinds of situations; the maximum transmission delays occurs 1) after the transmitted frames collide 16 times, 2) occurs when the backoff time has the greatest value, 2min (trial number 10), 3) occurs when D_{DET} has the greatest



value - twice the value of the transmission delay between the source stations and the destinations. In these situations, the theoretical maximum collision delays are calculated as the equation (1).

$$\begin{aligned}
 D_C &= \sum_{k=1}^{16} (D_{DET} + D_{JAM} + D_{BO \max}) \\
 &= 16(4D_{PROP} + D_{JAM}) + \sum_{k=1}^{16} 2^{\min(k,10)} t_s \\
 &= 4,185,920 t_b \approx 418.6 \text{ ms}
 \end{aligned} \tag{1}$$

where $D_{BO \max}$: Maximum backoff time with BEB algorithm

D_{JAM} : Jam transmission time

D_{PROP} : Propagation delay, D_{DET} : Collision detection time

From the above analysis, we can identify that the maximum transmission delay occurring in Ethernet is about 418.6 m s. This result shows that the direct application to industrial network is not concise because if the traffic of Ethernet increases, the number of collisions and the following transmission delays drastically increase.

Theoretical maximum transmission delays of switched Ethernet occur when N_q value is maximum which is determined by the network traffic and the quantity of switched buffer. If the network traffic is stable, and only the regular frames are generated, the value of N_q will be maximized because the processing ability of switches per hour is big enough when the frames with the same destinations are generated simultaneously and stored in switches. The maximum transmission delay time D_Q is like the equation (2).

$$\begin{aligned}
 D_Q &= \sum_{k=1}^{N_q} (D_{IF} + D_{TK}) = \sum_{k=1}^{N_q} (D_{IF} + \max(L_k + L_h, 576) t_b) \\
 &= N_q (96 + \max(K_k + 432, 576) t_b) \\
 &= 13 \times 672 \times t_b = 873.6 \mu \text{ sec}
 \end{aligned} \tag{2}$$

where D_{IF} : Interframe delay

D_{TK} : Message length of k^{th}

N_q : The number of frames stored in switches

L_k : Data size of k^{th} , L_h : Overhead of transmitted frames

According to equation (2), we know that if switched Ethernet is used, only a small transmission delay occurs. The reason for the result is that there is no collision in the switched Ethernet. The switched Ethernet satisfies the real-time requirements for the industrial networks, and it can be applied to railway signaling systems.



3 Network algorithm in railway signaling systems

In this paper, we study the methods of developing switched Ethernet reliability to apply them to railway signaling systems in which the safety is critical. Switched Ethernet satisfies the real-time requirements, but the reliability can not be guaranteed when each component is broken down because system components are related with each other such as Network-Interface Card (NIC), switch HUB, and communication lines. To solve this problem, we propose the multiple architecture where each component is duplicated, and estimate its performance.

3.1 Architecture of hardware redundancy

3.1.1 Switch HUB

In Ethernet communications, the data cables are shared, and all nodes have their individual addresses, and the communications are also done by the addresses through switching HUBs. In the beginning stage of switching HUB, unavoidable collisions occurred as traffics gradually increasing because the HUB, which never buffer total packets, has just a destination addresses and takes the way to transmit. Most of the currently realized switching HUBs don't bring any collisions excluding queue delay in switches, because they have independent queues of more than two in every port, and store the whole data in the way of Store & Forward, and then transmit the data after confirming destination addresses.

3.1.2 MultiPort Network Interface Card for fault-tolerant characteristic

Fault Tolerant is a system which is designed not to disturb system operations even in the case that some modules or components in the system have faults. The aim of the Fault Tolerant system is to continue operating the system regularly with no loss and destruction of data despite the occurrences of faults. The fault tolerance goes through the following three steps, these stages are proceeded and linked with running application software.

1. Fault Detection

Fault Detection is usually operated by Compare Logic constructed by hardware. When faults appear in systems, the decided modules or the systems come into the fault-condition. When the faults appear, OS analyzes every condition of hardware modules, and identifies which module makes faults.

2. Fault Diagnosis

The characteristics of faults are transient or hard. If the faults are hard, the modules of the systems are removed. If they are transient, that the self diagnoses result in no problem, the system first recognizes the fault as transient faults, and performs all operations.

3. Fault Recovery

Fault Recovery reconstructs systems by removing the modules which bring faults. One of the important functions, which keeps the fault tolerance, is to



duplicate data more than twice and then finally keeps the fault tolerance. In other words, when some data are created from application programs, the fault recovery always stores the data in two memories, and then transmits the data. Even when a module has some faulty data, fault recovery can keep the fault tolerance giving no influence on other systems. In this paper, we embody the Redundancy system by using NIC which has four Modules.

3.1.3 System architecture with fault tolerance

As mentioned above, to construct a fault tolerant system, multiple communication ports, and communication lines are needed. And also module redundancy should be embodied using multiple NIC, and redundancy in switches should be realized by linking switch HUBs doubly. If only one switch is installed, obstacles of total systems occur when faults are made in switches even though the redundancy in modules is guaranteed. Linking one communication line to each one port is another caution to avoid, because there is enough scope to make errors in the communication lines. The following Fig. 1 is the system architecture which meets those requirements.

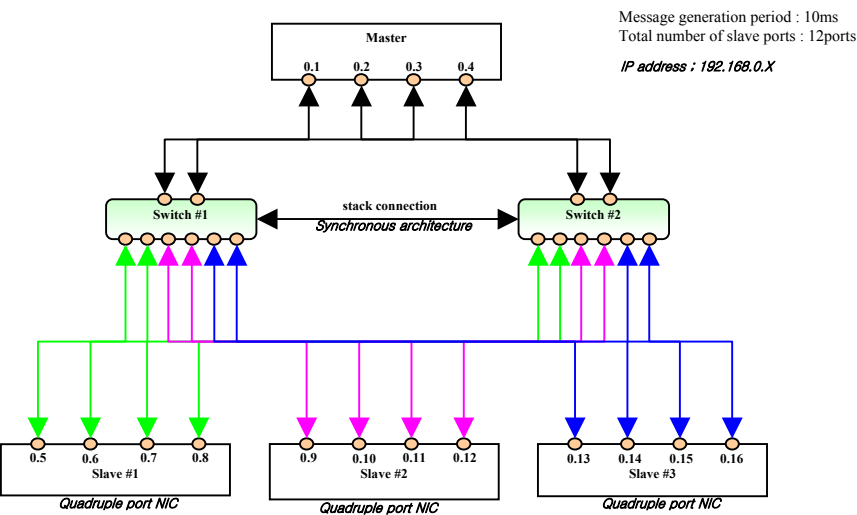


Figure 1: Overview of system architecture.

3.2 The development of fault detection and operation algorithm for communication nodes

In switched Ethernet, to double components, supporting software in application layer, 7th level of OSI reference is also necessary, not only a system framework of hardware. At first, more than two bindings of network cards are supported in one station for the duplication of network cards, and sockets and software buffers more than twice used for communication are needed.

The proposed system architecture is using four communication ports and four communication lines, and two switches, so the software for estimating the performance should be able to support system architecture as the above. Fig. 2 is the approximate program diagram of the software for full redundancy of the switched Ethernet. Explaining it approximately; a slave which generates messages continuously creates messages per 10 ms, and the message-generating routine creates an event of message generation to a slave message-sending routine when messages are generated. The slave message-sending routine which receives the event is constructed by the way of transmitting messages through individually bounding ports. The master that receives the messages plays a role to carry perfect data to application levels after copying the data entered individual received buffers and bring to voting routines, and comparing the data.

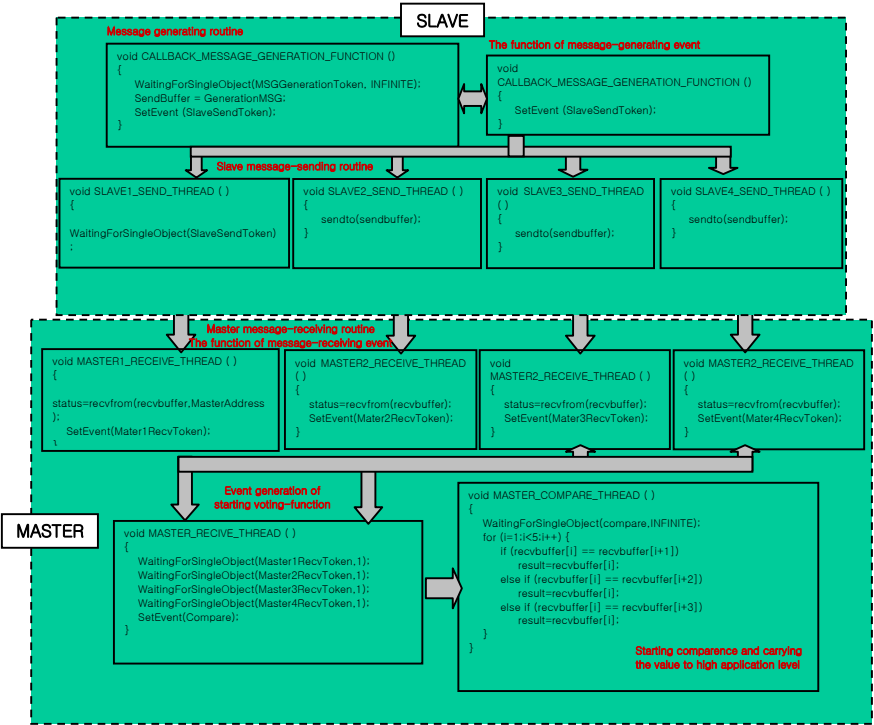


Figure 2: Test software configuration of proposed Ethernet architecture.

4 Conclusion

Railway signaling systems should necessarily guarantee reliability between the signaling systems because they are linked with other controlling equipment and construct bigger signal controlling systems, performing each function. The links of signaling systems has some problems with maintenance and repairs because the way of linking is of the point-to-point communication system.



The safety-critical systems, such as railway signaling systems, signify the reliability with emphasizing transmission delay. For these necessities, fault tolerance should be realized in the environment which requires high reliability when applying network like cars, automobiles, aviation vehicles. In this paper, we propose the reliable switched Ethernet architecture with redundancy systems and analyse the performing efficiency by using NIC with four modules.

We verified the proposed structures by experiments on bus redundancy for communication line reliability, module redundancy for confirming communication module reliability, and switch redundancy for switch HUB reliability. All error-recovery rates are 100% - all in the situational cases when communication circuits are cut during the communication between slaves and masters, when the uses of modules are stopped in the same conditions, and when the power supplies to switches are cut. The fault detection time is also less than 1ms. For these reasons, we consider that the proposed systems show the reliability enough to make the systems applied to railway signaling systems.

References

- [1] Hideo Nakamura, 'Development of Computerized Interlocking SMILENET', RTRI Report, 1986.
- [2] Nishida Shigenobu and etc., 'Replacement of Tokaido Shinkansen Signalling System Providing Advanced Train Controls in the 21st Century', KYOSAN CIRCULAR, Vol. 53, No. 2, 2002.
- [3] Kyung Chang Lee, Suk Lee, 'Performance Evaluation of Switched Ethernet for Real-time Industrial Communications', Computer Standards & Interfaces, Vol. 24, No. 5, pp. 411-423, Nov. 2002.
- [4] IEEE standard 802.3, Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications, 2000.



Section 13

Energy management

This page intentionally left blank

Information system for railway energy management

G. Hribar¹, B. Dremelj² & M. Sekavčnik³

¹*Genera Lynx d.o.o, Slovenia*

²*Slovenian Railways, Slovenia*

³*Faculty of Mechanical Engineering, University of Ljubljana, Slovenia*

Abstract

The energy market and railway transport have been opening up. One of the first European railway operators to respond to the changed conditions was the Holding Slovenske železnice (HSŽ, Slovenian Railways), constructing a comprehensive data reading and energy management system.

The presented system gives its users the ability to control the consumption on the site of supply and also on the site of usage. The system includes the supply grid and power substations, railway stations and other infrastructural objects, as well as energy usage readings from rail vehicles. Rail vehicles are equipped not only with energy meters but also with GPS and GPRS communication devices to transfer data into the energy management system in near-real time. This makes the equipment useful for billing and for lowering energy related costs. The data exchange with the traffic management system enables energy consumption forecasting and the optimisation of energy usage.

This article presents existing features and proposes the development of a complex railway energy management system. As a case study of collected data usage initial analyses on drivers' behaviour with the focus on coasting control and energy efficient driving strategy are presented.

Keywords: railways, information system, energy management, metering, on-board device, coasting.

1 Introduction

European electricity market has undergone structural changes in the past few years. Market liberalization, which now at least on paper becomes a fact for



bigger energy consumers in all EU members' states, brings more challenges than suppliers hoped, but less easy-saving- opportunities than consumers wished. In fact market prices in spite of established competition and freedom of choice after a short fall in the beginning now rise again. One reason is an increasing energy demand in the EU, another is a stricter emission regulation and the established emission allowances market, whose price per ton of CO₂ rose above all expectations. Although environmentally friendly, railways were hit by additional costs of emission trading more than other modes of transport.

To cope with those challenges energy consumers need much more information than in the era before the market opening. Cost reduction nowadays can be assured not only with lowering energy consumption, but also with efficient negotiation with the energy supplier for best supply conditions. For this the consumer needs to know the load profile of past and future energy consumption.

Improvement of energy efficiency is another way of lowering energy related costs. But this cannot be done without precise measurement and evaluation of present consumption, regarding all influence factors.

On the other hand the railway section itself faces great changes. Liberalization of railways services demands adaptation of existing service providers and also of infrastructure managers, which are not exposed to competition, but they need to be able to charge the network fees based on transparent and non-discriminatory rules to foreign operators. At the same time they should earn a reasonable return on its assets to assure adequate maintenance and development of railways infrastructure. Energy related costs are a significant cost factor of railways operation. In the case of an electrified railway network an infrastructure manager is responsible for the supply of energy. The infrastructure manager also must allocate energy consumption for each single train.

To face both challenges, rising prices of electricity and energy market opportunities on one side and liberalization of the railway section on the other, Slovenian Railways implemented an information system for energy management. The system collects data from all power substations for railway grid supply and bigger infrastructure consumers. Beside this, a pilot project of on-board electricity consumption and location metering was implemented and evaluated.

This article presents the development and implementation of the information system. A special focus is given to on board data collection and communication solutions. Another area covered in the article is evaluation of data collected and their usability for lowering of energy related costs.

2 Slovenian railways infrastructure and traction units

2.1 Infrastructure

Slovenian Railway lies on the crossroad of the 5th and 10th European transport corridor. Yearly volume of cargo traffic is 6.4 m gross ton kilometres and is



rising 5% per year. A large part of cargo traffic represents transportation from the port Koper to eastern and northern Europe.

Approximately 500 km of main lines are electrified. Power supply is assured through 17 substations, which transform electricity from the public electricity network to 3000V DC.

2.2 Engines

Most of the cargo trains on the main lines are hauled with engines type 342, type 362 and type 363.

Table 1: Engines properties.

		363	342	362
manufacturer		Alsthom	Asgen	Ansaldo
power (cont.)	kW	2750	1880	2640
power (one hour)	kW	2830	2280	3150
mass	ton	115	82	113
heating power	kW	990	1050	1050

All engines are supplied with 3000V DC voltage. The engines observed do not have possibility to recuperate energy.

Energy meters were installed as part of the pilot project and are mounted for a longer period so the installed devices did not have an influence on driver behaviour.

3 Measurements and data transfer

3.1 Network and infrastructure consumption

To retrieve data from the power substations on network and infrastructure consumption, the existing automatic meters reading (AMR) software Iskraemeco SEP2 is used. All power substations meters are equipped with phone modems. AMR software calls them once per day and collects data.

The network electricity usage profile is available in 15 minutes resolution.

3.2 Traction units consumption

3.2.1 Equipment

A data logging and communication device GSMBOX-R, produced by Genera Lynx, holds energy meter LEM EMT4, GPRS communication device GORLITZ Skalar and power supply. The device is mounted in the cabin of the engine.

Garmin GPS sensor with integrated embedded receiver and antenna is mounted on the roof of the engine.



3.2.2 Data collection and communication

Data collection from the energy meter is performed by regular reading of the serial port of the communication device. The created file in the communication device contains the profile of energy flow on the pantograph for active energy, reactive energy and recuperated energy in one minute resolution.

The Garmin GPS sensor feeds data of UTC time, latitude, longitude, hemisphere and speed over ground.

The communication device connects in regularly every few minutes interval to the private GPRS network. It sends data files through VPN connection into the application EGIDA, a product of Genera Lynx. EGIDA runs on a central server, collects the data and writes them into an SQL database. The application confirms receiving of the data and the data on board are flagged as sent.

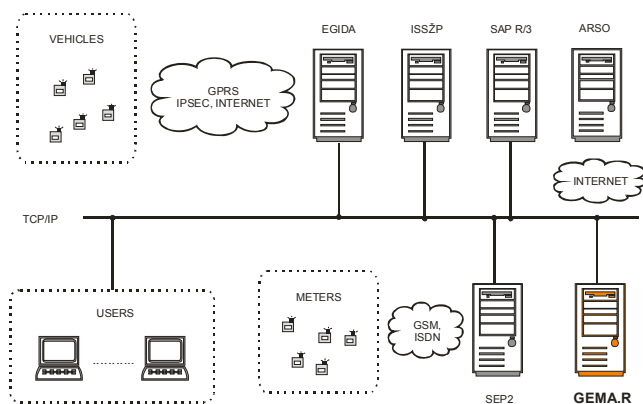


Figure 1: Communication block diagram.

The communication path is the same also when the engine leaves Slovenia. In this case roaming access to GPRS network is used.

Because the volume of data is low and speed of FTP connection high, transfer of data is performed typically in a few seconds. That is why the connection once established is almost always long enough to perform successful transfer.

If the connection cannot be established or if the connection falls during the data transfer, all data, unflagged as sent, are sent during the next successful connection.

To transfer data to GEMA.R EGIDA uses SOAP technology.

4 Information system for railway energy management

Genera Lynx's information system for energy management GEMA.R enables survey, analysing and forecasting of data that is relevant for energy consumption of railways traction and infrastructure units.



It is a web-based software solution using Java. Users access it through an enterprise intranet or the Internet, mobile GSM and PDA devices. It uses standard SQL database – MS SQL Server and OLAP technology for data mining.

4.1 Data sources

The energy data exchange server is only one source for the information system for railway energy management.

Besides energy usage on the engines GEMA.R collects daily read outs of substations and railways buildings meters from SEP2.

GEMA.R exchanges information also with traffic management system (ISSŽP) to get data of train properties (mass, length, type of the train and number of locomotives) and timetables for purposes of analysing and forecasting.

To perform billing and cost analyses from ERP system (SAP R/3) relevant data are imported.

Furthermore, for energy efficiency analyses and consumption forecasting actual weather data and weather forecasts are imported from the Slovenian environmental agency's (ARSO) FTP data server.

Manual input of data for consumption points without AMR equipment is possible. This is the case for smaller consumers, where data are updated once per month.

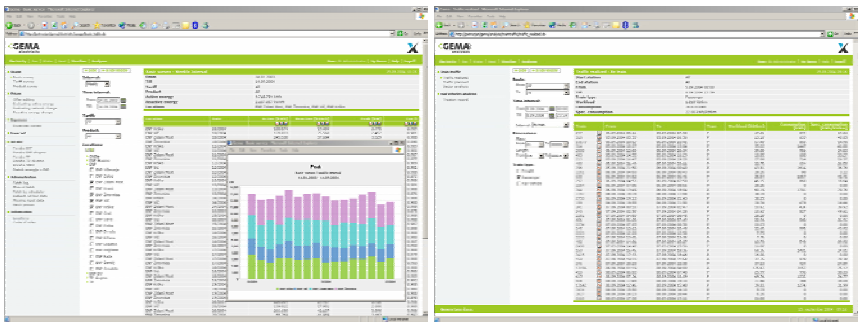


Figure 2: Screenshots of GEMA.R.

4.2 Functionalities

GEMA.R is based on the standard energy management system GEMA. It offers additional tools specifically developed for railways users.

Besides standard features of the energy management system GEMA (e.g. survey of energy consumption of any measurement point for chosen period and resolution, aggregation, price comparison, bill validation, etc.) within conjunction with TMS also total and specific energy consumption per each train or on daily, monthly and yearly basis are available. Apart from the train-based analyses also railway sections analyses are available with same features.

On the basis of timetables and energy related properties of archived journeys forecasts of energy consumption are calculated for a chosen train, for a particular substation and for all consumers together.



Cargo train timetables are de-facto not fixed up to the minute and usually can be adapted in order to reduce energy costs.

Energy consumption forecasts are valuable data for lowering peak power demand on each substation. Besides these costs could be reduced with improved load profile (raising lowest and lowering highest 15 minutes or 1 hour consumption in a day) and to minimize the difference between the bought volume and consumed volume of electricity.

5 Analysis of sample data

5.1 Background

The aim of the analyses of the sample data from two characteristic sections was to get additional guidelines for further development of tools of the information system for an energy management system. For the purposes of energy management of Slovenian railways relevant factors for an optimal driving strategy on those sections was researched.

Regarding the existing literature coasting control could improve energy efficiency (Wong and Ho [1], Hwang Hee-Soo [2]). Regarding to Albrecht [3] the optimal driving regime consists of a maximum of four phases (driving with maximal acceleration, travelling with constant speed, coasting and operational braking to target with maximal deceleration). But as seen in Lukaszevicz [4] driver behaviour in the real world differs much from optimum driving strategy. An optimal compromise between trip time and energy consumption was researched with many tools [3, 2]; the study of Rongfang and Golovitcher [5] brings a comprehensive review of solving the problem of energy efficient train control. Those studies rely mostly on simulations and are not often supported by extensive experimental data.

In this work the primary focus is given to experimental determination of the influence of the degree of coasting on energy efficiency.

5.2 Chosen sections

Sample data form two sections, very different in topology but both with heavy cargo traffic, were observed; first from the Koper cargo station to Divača and another from the central cargo station Ljubljana Zalog to Zidani most. From Koper station to Divača a long steep hill occurs (25‰) while the track form Ljubljana Zalog to Zidani most is characterised by a gentle decline.

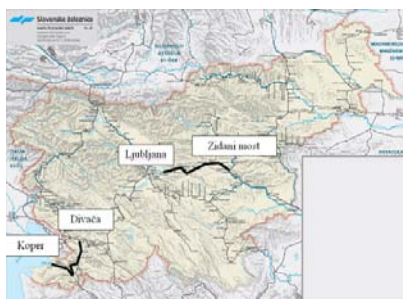


Figure 3: Slovenian railway network and sections chosen.



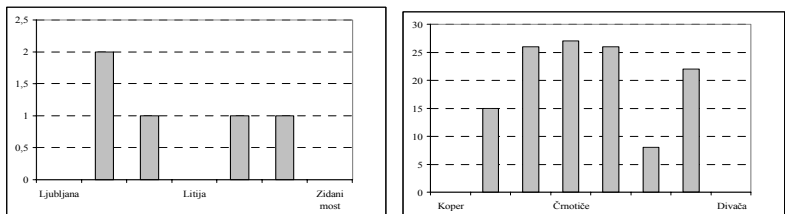


Figure 4: Characteristic resistance of the tracks (daN/t).

Because of the low resistance on the section Ljubljana Zalog - Zidani most one engine hauls a train with a mass up to 2000 tons. From Koper cargo station to Divača one engine is enough only for hauling trains lighter than 725 tons. When hauling heavier trains an engine with an on board meter device was combined with non-metered engines and complete energy consumption of the train was not available. That is why on this section only trains up to 725 tons were included in the analyses.

5.3 Period and volume of measurement

Data was collected from 1st May 2005 to 31st January 2006.

During that period more than 1000 trains with meters on board the engine drove on the observed sections. To assure comparable data only trains which entered the section in Ljubljana (Koper) and left the section in Zidani most (Divača) were included in the analyses. Regarding the above mentioned limitations this lowers the number of trains included to approximately 100 (10) trains.

5.4 Influence of coasting ratio on specific consumption

Coasting time is defined as the time of journey when power is equal to or lower than 1 kWh / min and the speed is higher than 3 m/s.

$$\text{coasting ratio} = \frac{\text{coasting time}}{\text{total journey time}} \quad (1)$$

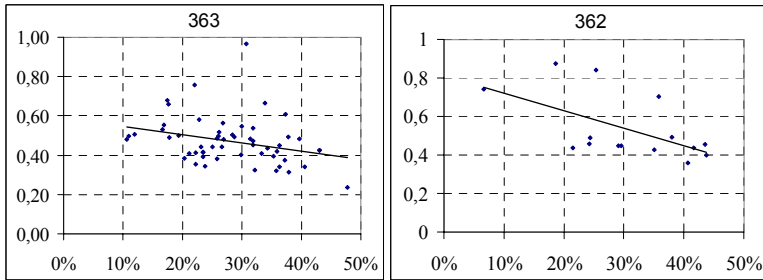


Figure 5: Specific consumption (kWh/ton) as a function of a coasting ratio (%) on section Ljubljana – Zidani most for engines 363 and 362.



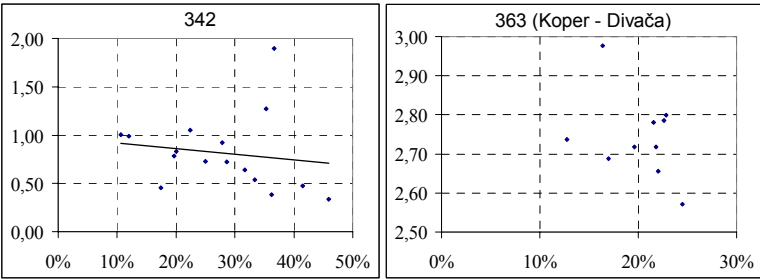


Figure 6: Specific consumption (kWh/ton) as a function of a coasting ratio (%) on section Ljubljana – Zidani most for engine 342 and on section Koper – Divača for engine 363.

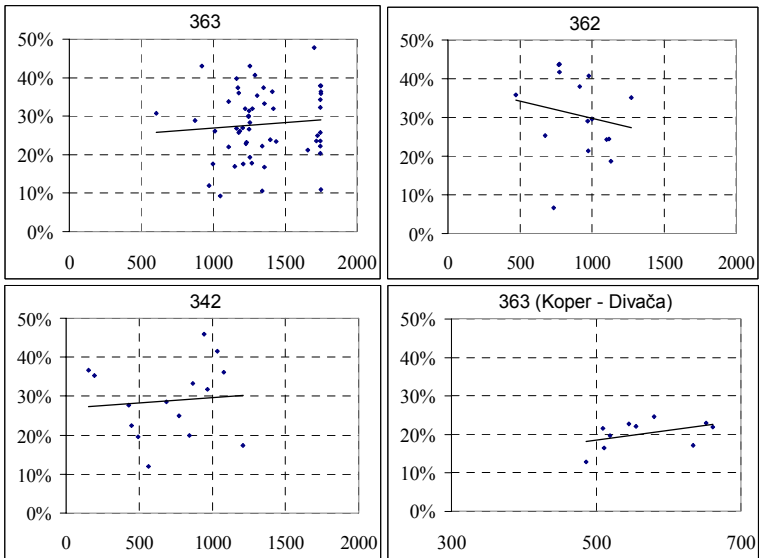


Figure 7: Ratio of coasting (%) as mass (tons) function on sections Ljubljana – Zidani most and Koper – Divača.

5.5 Influence of mass on coasting ratio

While there are no tools for coasting point definition available for the drivers, the decision depends only on their experience. There is no evidence that drivers of heavier trains on either of the observed sections use their accumulated kinetic energy for a longer coasting period.

5.6 Driving strategy of efficient and inefficient journey

The driving strategy of two similar trains with significantly different consumptions are observed to define the characteristics of efficient driving.

The journeys of both trains were performed between Ljubljana and Zidani most. Considering similar train properties, table 2, it may be expected that the reason for a difference in consumption lies in different driving strategies.

Table 2: Trains characteristics.

train id	speed (m/s)	consum. (kWh)	length (m)	mass (ton)	coasting ratio	$\Sigma \Delta speed$ (m/s)	specific consum. (kWh/ton)
7551	13.5	632	263	1253	28%	151	0.50
7579	13.6	533	260	1254	43%	144	0.43

$$\Sigma \Delta speed = \sum_n abs (speed_i - speed_{i-1}) \tag{2}$$

where speed_i the is speed of the train in ith minute.

Train 7551 has significantly lower coasting ratio. This indicates longer period of powering and more intensive braking. A more dynamic journey is indicated also with higher sum of $\Delta speed$, but here the difference is low.

The coasting pattern for both trains is similar, fig. 7, regarding the ends of the coasting periods. But train 7579 usually starts coasting earlier then train 7551 and the average coasting period is longer.

Considering similar properties of the train and a similar level of acceleration (fig. 8, table 2), which is – because recuperation of energy is not possible - directly correlated to the energy needed, it could be assumed that the main reason for lower energy consumption of train 7579 is a higher coasting ratio. This higher ratio is achieved mainly by earlier starts of coasting periods.

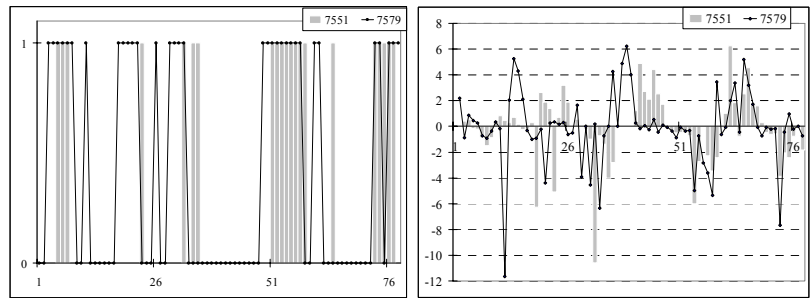


Figure 8: Coasting periods (when value = 1) and acceleration ($\Delta m/s$) as a function of time (min) for trains 7551 and 7579.

6 Conclusions

An information system for energy management, implemented in Slovenian Railways, enables the reduction of energy related costs with lower price and lower energy usage. With a better knowledge of the load profile of the substations Railways improved their position on the energy market and lowered average price.



With a comprehensive database of engines energy usage with simultaneous location and speed data and supplemented with information from the traffic management system extensive analyses of energy efficiency of a particular engine or the whole railway network are possible.

On the basis of data collected the traffic on the two sections has been examined here. The topology of the track limits coasting, but on both sections possibilities for further improvements of energy efficiency with longer coasting ratio have been shown.

Furthermore, two journeys with similar train properties but different energy consumption have been reviewed in detail with respect to driving strategies. The starting point of coasting has an important influence on energy consumption.

References

- [1] K.K. Wong, T.K. Ho, Dynamic coast control of train movement with genetic algorithm, *International Journal of System Science*, Vol. 35, No. 13-14, pp. 835 – 846, 2004.
- [2] Hwang Hee-Soo, Control Strategy for Optimal Compromise Between Trip Time and Energy Consumption in a High-Speed Railway, *IEEE transactions on systems, man, cybernetics – part A: systems and humans*, vol. 28, No. 6, pp. 791 – 802, 1998.
- [3] T. Albrecht, Reducing power peaks and energy consumption in rail transit systems by simultaneous train running time control, *Proc. of Computers in Railways IX: Computer Aided Design, Manufacture and Operation in the Railway and Other Advanced Mass Transit Systems (COMPRAIL 2004)*, eds. J. Allan, C.A. Brebbia, R.J. Hill G. Sciutto & S. Sone, WIT Press: Southampton, Boston, pp. 885 – 895, 2004.
- [4] Lukaszewicz P., *Energy consumption and running time for trains*, Royal Institute of Technology, Stockholm, 2001.
- [5] Rongfang Liu R., Golovitcher I. M.: *Energy-efficient operation of rail vehicles*, Transportation Research Part A 37, pp. 917 – 932, 2003.



Impact of train model variables on simulated energy usage and journey time

P. Lukaszewicz

Aeronautical and Vehicle Engineering, KTH, Stockholm, Sweden

Abstract

Several train model input variables, such as running resistance, line voltage, adhesion, braking release time and braking gain time, are studied. An analysis is performed on how variations in the variables impact relatively on calculated energy usage and running time of trains. The study shows that for the calculation of energy usage the simulations are most sensitive to variations in running resistance, followed by line voltage, adhesion, braking release time and braking gain time. For the running time, the study shows that variation in mechanical rolling resistance and air drag has a relatively small influence provided that the tractive force is big enough. If the line voltage and adhesion, which affect here the tractive force, drop below certain levels the running time increases dramatically. The braking release and gain times have little influence on the running time. The results also show which variables should be paid extra attention to, when constructing a train model.

Keywords: train modelling, train data, sensitivity, power consumption, energy usage, running time, simulations, ERTS.

1 Introduction

The correctness of computed results of energy usage and running time of trains in a railway network is dependent upon the chosen train model and input data. Therefore it is of interest to examine quantitatively how much the results can differ from each other if the input data used by the same train model varies and which data should be paid extra attention to.

By means of sensitivity analysis, the impact of the following variables is studied for a SJ Rc4 loco hauled freight train:



- *Running resistance*, which is the total force acting against the travel direction.
- *Adhesion*.
- *Tractive force* (due to variation in catenary voltage).
- *Braking gain time*, which is the time it takes to obtain the desired braking force, from when the driver starts braking.
- *Braking release time*, which is the time it takes to reduce the braking force to zero, from when the driver stops braking.

Section 2 describes the method and models. The results are presented in section 3 and are discussed in section 4.

2 Method and models

This sensitivity analysis on how variation in input data affects the final results on computed energy usage and running time is here performed by means of the Energy and Running Time Simulator, ERTS. ERTS is a simulation program developed by KTH and has verified models and data, versus full-scale measurements, of trains and drivers. The verification shows that the discrepancy between calculated and measured train energy usage is within the measurement error of approx. 2% [1].

The train models are detailed especially with respect to braking and tractive forces, electrical efficiency, running resistance, adhesion and slippage.

The driver models in ERTS are developed from full-scale measurements [2]. Observations were made on how real drivers are handling the trains especially with respect to track profile, signalling and type of train and service. The developed driver models, not included here, can drive a train as an average driver would drive, or drive in an optimised way with respect to energy usage or running time.

The driver model in this study is constant and set to drive the train strictly in accordance with the signalled speed. The acceleration is performed at maximal powering. Braking is performed as late as possible with respect to the braking ability which is set to 1/3 of the maximal braking force of the train. This level of the braking ability is obtained from observations on how the trains are driven in reality. The models are described in [1].

2.1 The train model

The train model represents a loco hauled freight train of mixed consist.

The locomotive is of type SJ Rc4 and the tractive force diagram for two different catenary voltages and powering levels is shown in Figure 1, together with the tractive force limit, F_{α} , due to adhesion as it is modelled in ERTS.

The calculated magnitude of the tractive force, F_w , takes into account the powering level, effect of speed, catenary voltage and the tractive force, F_{α} , available with respect to adhesion. In this study, no wheel slippage is present.



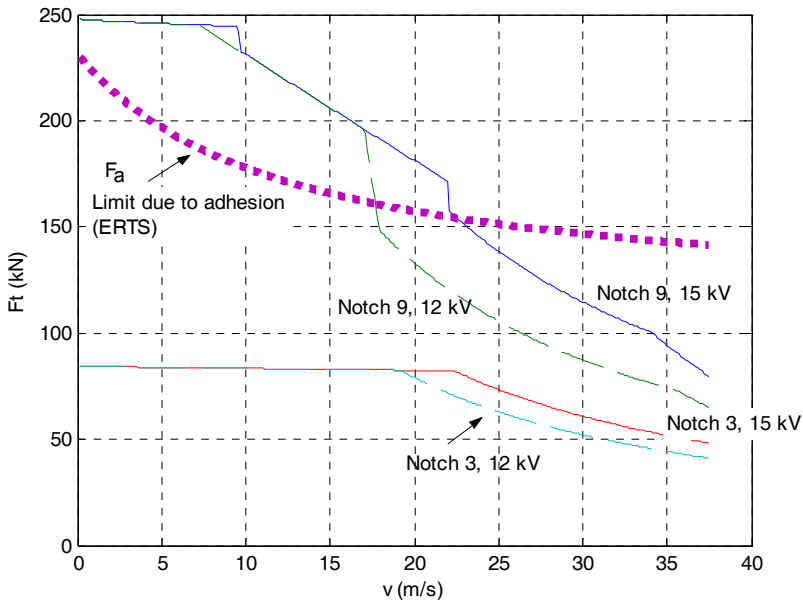


Figure 1: Tractive force diagram. Notch 9 is the maximal powering level.

This means that the train speed is the same as the tangential speed at the peripheral of the wheels of the locomotive. The tractive force at the wheels, is calculated by:

$$F_w = \min(F_t, F_a) \quad (1)$$

The total energy usage, of the train is calculated at the pantograph level for two cases; E_1 , when a tractive force is present and the train is moving, and E_2 when the train is coasting, braking or not moving.

$$E_{tot} = \left\{ \begin{array}{l} E_1 = \sum_{i=1}^n \frac{(F_{w(i)} + K_j a_i) v_i (1 + \zeta) \Delta t_i}{\eta(p, v)_i} \frac{1}{3.6 \cdot 10^6}; \quad v > 0, \quad F_w > 0 \\ E_2 = \sum_{i=1}^n P_{0(i)} \Delta t_i; \quad v = 0 \text{ or } F_w = 0 \end{array} \right\} = E_1 + E_2 \quad (2)$$

where, E_{tot} is total energy usage in kWh, n is the total number of time steps Δt during a simulation. K is a constant accounting for the rotational masses, a is the acceleration, ζ is the slippage ($=0$), η is the efficiency of the locomotive as a



function of power, p , and speed v , and P_0 is originating from the auxilliary power. The total running time is calculated from

$$T = \sum_{n=1}^i \Delta t_i \text{ (s), for } v > 0 \tag{3}$$

The freight wagons in the train set have 2 axles/wagon and are of two types; open type Oms and covered type Hbis. Basic data for the test train is shown in Table 1:

Table 1: Nominal and basic data for the test train.

Length, incl. loco	418.5 m
Mass, gross incl. loco	1197 t
Mass of locomotive SJ Rc4	79 t
Axles, trailing	52
Max speed	100 km/h, 27.8 m/s
Axle load, average	21.5 t
Braking gain time, nominal	15 s
Braking release time, nominal	30 s
Braking level used	1/3 of max

The reason for choosing this train configuration is because of the existence of measured data [1] on energy usage, running resistance, tractive force, efficiency, braking ability and time lags in the tractive and braking systems.

2.2 Track model

The track model represents a tangent CWR. The length of the track is 88 km. A simulation with nominal input data for the train model results in a running time of 3597 s. The signalled speed restrictions are according to Table 2:

Table 2: Speed restrictions for the track model.

Distance (m)	speed (km/h)
0	100
20490	40
21364	100
38152	70
39288	100
44106	70
44566	100
51322	40
52534	100
88000	100



3 Impact of variables on energy usage and running time

3.1 Simulation with nominal input data

Figure 2 shows the speed profile for the train obtained from simulation with nominal input data. Table 3 shows the numerical results. This is the reference case, with which all other results are compared with in this study.

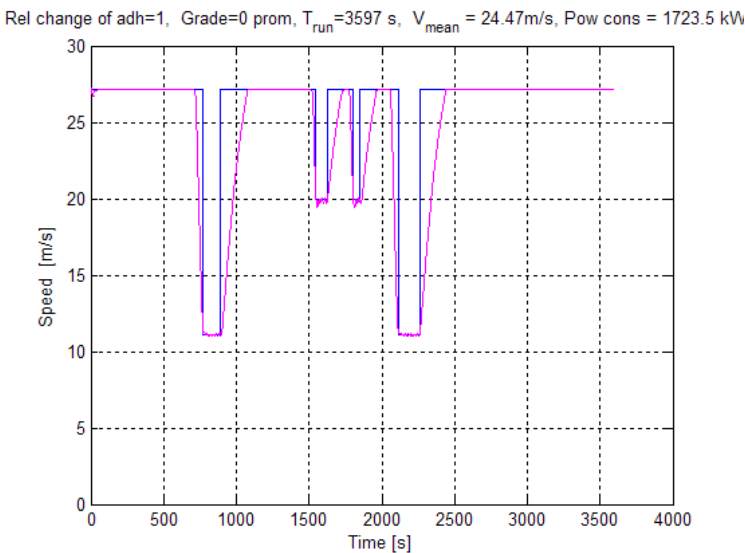


Figure 2: Speed profile from simulation with nominal input data.

Table 3: Results from simulation with nominal input data.

Constant grade (‰)	E_{tot} (kWh)	T (s)	Mean speed (m/s)
0	1723.5	3597	24.47
5	3329.9	3758	23.42

3.2 Running resistance

The nominal running resistance, F_{R0} , of the train set is obtained from full-scale measurements [1] and is calculated as a function of speed, v , by:

$$F_{R0} = 11961 + 229.1v + 41.4v^2 \quad (4)$$

The impact of variation of running resistance on energy usage and running time is shown in Figure 3.



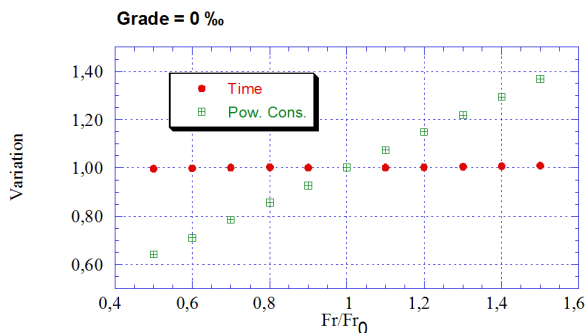


Figure 3: Impact of variation of running resistance on energy usage and running time.

In this case, the impact on running time is small, but big on the energy usage. If the resistance has large errors from input data together with resistance originating from grades, the tractive force of the locomotive might not be sufficient. In this case severe delays will be present.

3.3 Adhesion

The available nominal adhesion is calculated in ERTS by the Curtius-Kniffler formula [3] which has been modified [1] to better suit full-scale test data.

$$\alpha_0 = 0.9 \left(\frac{7.5}{44 + 3.6v} + 0.161 \right) \tag{5}$$

The results are shown in Figure 4. If the adhesion is higher than nominal, almost no variation occurs. However, if the adhesion ratio for this case starts decreasing below approx 0.7, the running time starts increasing due to insufficient tractive power limited by the adhesion. Energy usage decreases mainly because of lower average speed which reduces the aerodynamic drag.

3.4 Line voltage

The tractive force of the locomotive SJ Rc4 is affected by the line voltage, see Figure 1. A voltage drop decreases the tractive force from the train speed of 17 m/s and up.

The variation of running time and energy usage due to variation of line voltage is shown in Figure 5. The nominal voltage is 15 kV.

3.5 Braking gain time

The variation of braking gain time has for this studied case very small impact on the running time and energy usage, as shown in Figure 6.

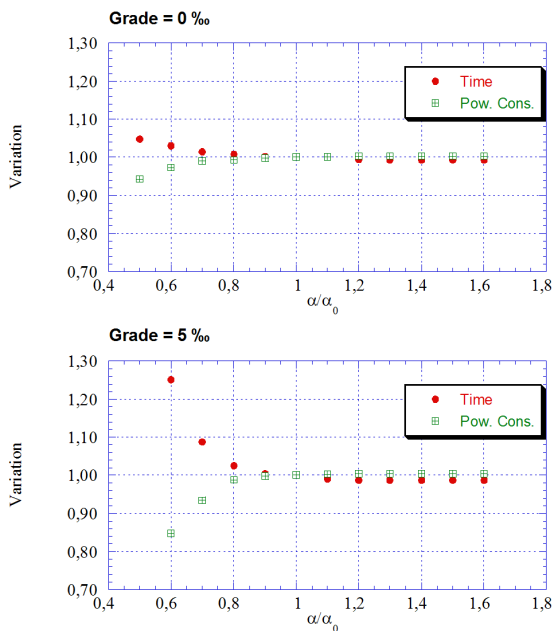


Figure 4: Impact of adhesion on energy usage and running time for grade 0 and 5‰.

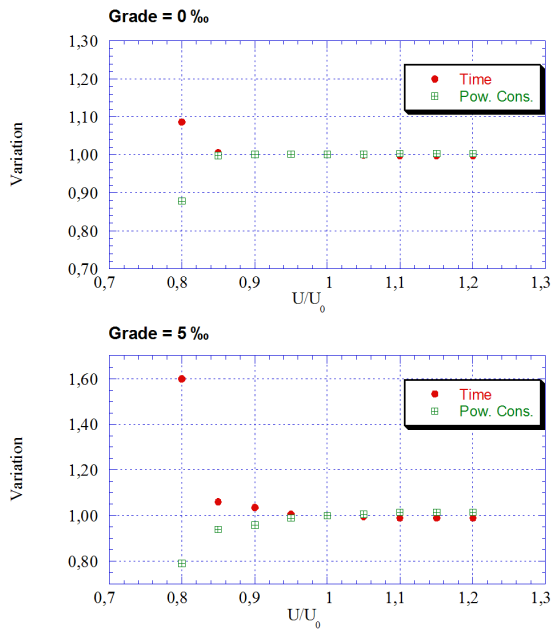


Figure 5: Variation of energy usage and running time due to variation of line voltage.



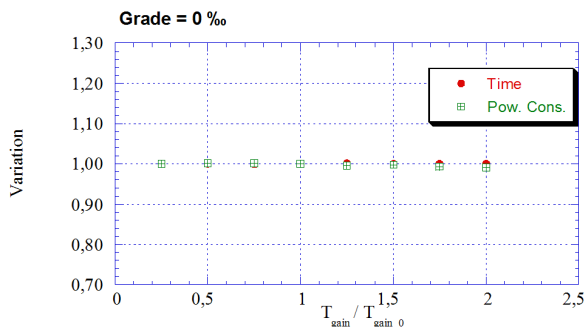


Figure 6: Variation of running time and energy usage due to variation of braking gain time.

3.6 Braking release time

The variation of braking release time has a slight impact on energy usage. If the braking release time is reduced, compared with the nominal 30 s, a decrease in energy usage is distinguished, Figure 7.

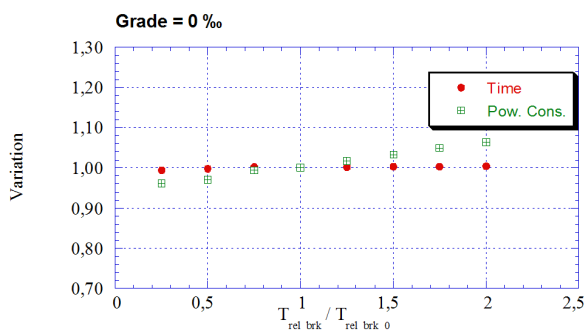


Figure 7: Variation of energy usage and running time due to variation of braking release time.

4 Conclusions

This study shows in a quantitative way the importance of choosing correct input data and their significance. It is therefore important to have up to date models, to collect train data, maintain databases and to have information on how and for which circumstances the data should be used.

Variation of *running resistance* has little effect on running time, provided the tractive force is sufficient. The energy usage is strongly dependent upon the running resistance.

When the available *adhesion*, as modelled in ERTS, drops under a certain level the energy usage drops as well. The running time increases significantly.

When the *line voltage* drops and the tractive force is not sufficient, the energy usage drops as well. The running time increases significantly.

Variation of the *braking gain and release times* showed little significance in this study.

In this study, only the train model data is studied. An another important factor is the driver behaviour which has a strong impact on energy usage .

References

- [1] Lukaszewicz P., *Energy Consumption and Running Time for train*. KTH Stockholm 2001. TRITA-FKT 2001:25. ISSN1103-470X.
- [2] Lukaszewicz P., *Driving describing parameters, energy consumption and running time*. Computers in Railways VIII. Comprail 2002 Lemnos.
- [3] Andersson, E. Berg, M., *Railway systems and vehicles* (in Swedish). KTH



This page intentionally left blank

Section 14

Power supply

This page intentionally left blank

Investigation into the computational techniques of power system modelling for a DC railway

A. Finlayson¹, C. J. Goodman² & R. D. White¹

¹*Atkins Rail, UK*

²*University of Birmingham, UK*

Abstract

The use of computer simulation techniques is now a fundamental part of the design process for electrified railways and at the feasibility stage clients will often request detailed calculations to be performed for the basic design. This will establish a level of confidence in both the project and basic design parameters that will justify the capital expenditure further on in the project life-cycle.

This paper will address how the total impedance of a railway network may be represented, where the impedances of the traction return circuit and traction power system are either combined together to form one impedance or studied independent of one another. The accuracy of the modelling in this manner, particularly how it affects the accuracy of rail voltage results shall be assessed. It will also examine how much impedance is typically in the rails and what proportion this needs to be before it has an unacceptable effect on the numerical results. To assess this, a range of proportions will be studied, for example 70% of the total impedance to be modelled in the conductor with 30% modelled in the rails, 60%/40%, etc. In this way, a proper scientific assessment of the combined or split calculation methods can be made via simplified case studies.

Keywords: DC railway, modelling, simulation, computational techniques, accuracy, numerical method, rail voltage, stray current.

1 Introduction

DC electrified railway systems across the world are growing, both in passenger and freight traffic, as an alternative solution to increasing road congestion. DC electrified railways account for approximately 50% of electrified railway lines throughout the world [1]. This growth may be in the form of new build railway



systems or the upgrade of existing railway systems and is inclusive of metro, tramway, light rail and heavy rail systems. However this growth has to be managed to ensure that the proposals for an improved infrastructure are designed both safely and correctly and that the design delivers value for money. Computer simulation is the tool that many consultants and manufacturers, associated with the rail industry, are using to evaluate proposals and validate final designs prior to the launch of the project and any major capital expenditure. The use of computer simulation techniques is now a fundamental part of the design process for electrified railways and at the tender stage clients will often request detailed calculations to be performed for the basic design [2]. This will establish a level of confidence in both the project and basic design parameters that will justify the capital expenditure further on in the project life-cycle.

1.1 DC railway modelling and simulation

Modelling and simulation are the names we use for applying the laws of physics and logic via the processing power of computers to predict the behaviour and performance of railway infrastructure [2]. The various levels of engineering modelling and simulation with respect to the DC railway may be considered as; timetable and operational planning, train performance and signalling, traction equipment and the power supply system. These engineering levels are not considered to be exhaustive but what is important is the power system supply, as this paper is primarily concerned with the modelling of the power system for the DC railway.

2 The DC railway

2.1 Operational functions

The DC railway usually has the primary operational function of passenger transportation, whereas the AC railway has the dual operational functions of passenger and freight transportation. However the DC railway may also be used within the mining industry, as well as freight transportation [3], so it may be considered to have dual operational functions also. However, for the purpose of this paper the DC railway shall be considered as predominantly suburban railways often referred to as transit systems. Transit systems may be classified based upon the demand for traction current from the power system. For example, a metropolitan railway (metro) may be considered to be a mass transit system as there will be more trains carrying more passengers with short headways between trains and therefore a large demand for traction current from the power system. Conversely, a tram railway may be considered to be a light rapid transit (LRT) system as there will be less trams and also less demand for traction current from the power system than the mass transit system.

Either of these systems may utilise an overhead line or third/fourth rail feeding arrangement. The following DC traction supplies are commonly used [4]:



- 3000 V (overhead line)
- 1500 V (overhead line)
- 750 V (third rail and overhead line)
- 630 V (third and fourth rail)

2.2 Components of the DC railway

The DC electrified railway comprises of many interacting variables or components. These variables are usually contained within sub-systems such as; civil engineering, permanent way, electrification, signalling, tele-communications, rolling stock, station services and third parties outside the railway environment. Whilst this list is typical (non-exhaustive), all of these variables must be considered both individually and interacting with one another during the design process of a railway system. However, for the purpose of this paper the components that are of most concern are;

- Electrification
- Rolling stock
- Signalling

2.3 Other considerations for modelling

In studying the DC railway power system, there are two further technical problems that need to be included in the circuit model of the DC power system. Firstly, there is the variation of rail voltage with respect to the train operational timetable and secondly there is the magnitude of stray current with respect to the train operational timetable.

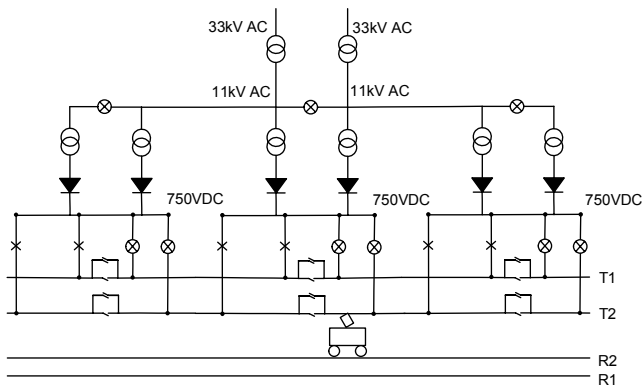


Figure 1: The DC railway represented in a single line diagram.

2.4 DC railway operation

Figure 1 shows a typical two track DC railway with one feeder station with a primary voltage of 33 kV AC and a secondary voltage of 11 kV AC. The



11 kV AC is then rectified and an output voltage of 750 V DC is produced which in turn will power the rolling stock on the rail network. The rolling stock will draw traction current from T2 and will return traction current through R2. However not all of the traction return current will return through the running rails and this may be due to the earthing and track bonding arrangements associated with the DC railway.

3 Choosing the model and method

3.1 The basic model

There are many methods available to us in solving the power system model of a DC railway. Examples of these may include multi-conductor modelling, finite element analysis or transmission line theory to name but a few. However, given the nature of the DC railway, the power system model may be considered to be dynamic, i.e. the electrical load changes with demand and with multiple electrical loads, i.e. trains. Whilst these parameters may appear complex, the DC railway power system is easier to model than the alternative AC railway. This is because the effects of capacitive and inductive reactance in the equivalent steady state circuits may be ignored due to the general principles of DC circuit theory [5]. Applying this to figure 1 will give rise to the equivalent circuit shown in figure 2, where Z_{c1} , Z_{c2} and Z_{c3} represent the impedance of the traction power conductor, whilst Z_{r1} , Z_{r2} and Z_{r3} represent the impedance of the traction return conductor.

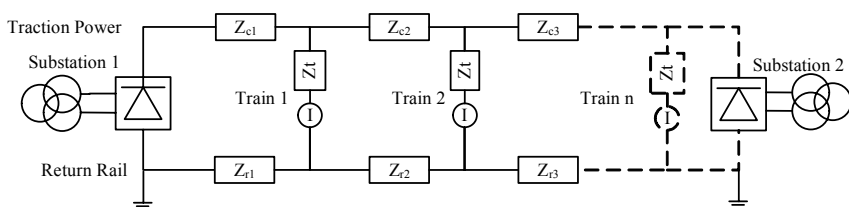


Figure 2: Electrical circuit for DC railway.

Figure 2 may still not be considered to be the simplest model of the DC railway. The electrical substations may be modelled in their Thévenin equivalent model, a voltage source (V_s) with a source impedance (Z_s) and the train may be modelled as an impedance (Z_t) that is dependent upon whether the train is motoring, coasting or braking.

Figure 3 now shows two cases of the simplified circuit, with the Thévenin equivalent voltage source representing two electrical substations and a single train represented by an impedance somewhere between the two electrical substations. The circuit shown to the left of figure 3 represents the DC railway where the traction power and traction return conductors are split from one



another, whilst the circuit to the right shows them combined. The circuit to the right suggests that the DC railway power system model may be solved as a simple DC circuit, using short transmission line theory [6].

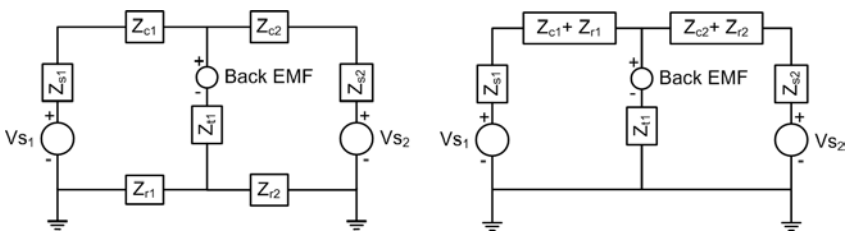


Figure 3: Simplified electrical circuit of a DC railway.

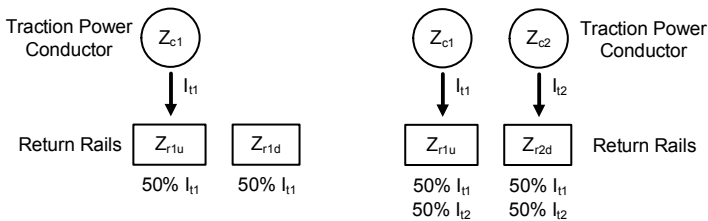


Figure 4: Developing the combined impedance model.

3.2 Representing the traction return conductor in the combined impedance model

Figure 3 is representative of a one track railway. In practice, the DC railway, under consideration for this paper, will usually be a minimum of two tracks, namely up (u) and down (d). Considering figure 4, there are two trains present, one on each track and there are four traction return conductors, or as is more commonly known return paths. These return paths are considered to have identical impedance characteristics and the percentage of traction return current (I_{t1} and I_{t2}) in each return path is shown.

Considering the configuration shown to the left of figure 4, will give rise to the equivalent impedance of all the return paths combined being expressed as:

$$Zr_n = \frac{Zr_{1u}}{n} \quad \Omega \tag{1}$$

where ‘n’ is equal to the total number of traction return conductors in the circuit. However, considering the configuration to the right of figure 4, the equivalent impedance of all the return paths combined will be expressed as:

$$Zr_n = \frac{T * Zr_{1u}}{n} \quad \Omega \tag{2}$$

where T is the number of trains present. However, because the number of trains present between electrical substations and on different tracks varies, this will be difficult to program in a computer simulation tool. The equivalent impedance of all the return paths combined may be expressed as an alternative to equation 1 by

$$Zr_n = Zr_{lu} \quad \Omega \tag{3}$$

This paper will assess which the accuracy of modelling the equivalent return rail impedance as either equations 1 or 3 for a number of tracks, as these are feasible methods that may be used by computer simulation tools.

3.3 The method

Figure 3 may be solved via a number of well known network analysis techniques such as mesh or nodal analysis [5]. For simple circuits such as that shown in figure 3 where there is only 1 train present, the solution may be solved by deriving linear equations and solving them as simultaneous equations. If the circuit contains more electrical loads, it is better to solve the circuit via a series of linear equations represented in matrices where the general equation may be given as

$$[V] = [Z] * [I] \tag{4}$$

The computational efficiency of the numerical procedures undertaken by both mesh and nodal analysis may be measured by counting the number of numerical operations (OC) i.e. multiplication, division, addition and subtraction for each equation ‘N’ to be solved, i.e. ‘N’ is equal to the number of nodes in the circuit or the number of loop currents in the circuit, dependent upon the network analysis technique chosen. This may be defined as [10];

$$OC = \frac{2N^3}{3} + \frac{3N^2}{2} - \frac{7N}{6} \tag{5}$$

Table 1 shows the difference in operational counts taking into account the simplified models shown in figure 3. Table 1 illustrates that nodal analysis has a much lower OC when 4 or more linear equations are required to be solved. This can be attributed to the fact that in nodal analysis, each node represents an equation and one node has to be set to zero.

Table 1: Summary of OC.

Number of Traction Return Paths	Mesh Analysis		Nodal Analysis	
	N	OC	N	OC
1	2	9	2	9
2	4	62	3	28
4	8	428	5	115
6	12	1354	7	294
8	16	3096	9	527

If using mesh analysis, each loop current represents an equation and each and every loop current must be used. This means that for an equivalent circuit that requires 4 or more equations to be solved, mesh analysis will always have one more equation to be solved than nodal analysis, thus meaning that nodal analysis is the more efficient method.

3.4 Injecting train currents

If the combined impedance circuit of figure 3 is used to solve the power system then it presents us with a problem as the technical issues cited in section 2.3 are not represented in the model, whereas the rail voltage is represented and calculated directly in the split impedance model. This problem may be overcome by either including these parameters in the model for real time simulation or they can be accounted for in a post processing program, i.e. a separate program that operates on the results from the initial simulation [7]. This may be done by injecting the train currents (I_t) into a homogenous rail. The ground impedance (Z_e) is represented as a lumped parameter that ignores the effect of ground capacitance, because any reactive component represented by the ground will be so small that it can be ignored [4, 8, and 9] and in any case is only relevant for transient conditions, as is shown in figure 5. The current flowing in the earth (I_e) represents the stray current and this may also be studied as part of the post processing program.

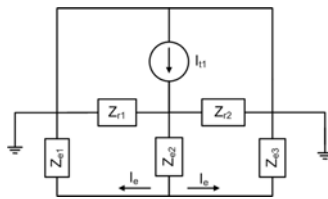


Figure 5: Post process injecting train currents model.

4 Case studies

The theoretical analysis of railway modelling techniques discussed in section 3 may be applied to the practical railway, where numerical analysis can be applied to ascertain the accuracy of modelling between the split impedance method (with the traction return path impedance modelled as per equation 2) and the combined impedance method with the equivalent traction return path impedance modelled as per equations 1 and 3. The network analysis technique chosen was nodal analysis as this calculates the train voltage directly and is considered to be more efficient than mesh analysis, as defined in section 3.4. The following railway types have been identified as case studies;

- 1500 V DC Overhead Line Railway (OHL)
- 750 V DC 3rd Rail Conductor Railway
- 750 V DC Overhead Line Railway



Practical traction return circuits, traction power system and voltage source impedances have been used in the case studies. A range of proportions have been studied, for example 60% of the total impedance to be modelled in the traction conductor (α) with 40% modelled in the rails etc., where the summation of the traction powers system and return circuit impedances is equivalent to the loop impedance Z_L . Each case study includes six traction return circuit with up to six trains present, i.e. one on each track.

4.1 Calculating the error

The train voltage may be expressed as a percentage of the supply voltage derived by the electrical substation. This may be expressed generally as;

$$V_t = \frac{V_{train}}{V_{S1}} \times 100 \quad \% \quad (6)$$

where V_{train} is the train voltage and V_{S1} is the electrical substation voltage. Equation 6 will satisfy both the split and combined impedance analyses. The error in train voltage calculation may be expressed as;

$$\%error = V_{t_{split}} - V_{t_{combined}} \quad \% \quad (7)$$

where $V_{t_{split}}$ is the result of equation 6 for the split impedance model and $V_{t_{combined}}$ is the result for the combined impedance method. A back calculation may be performed using the result of equation 7 to calculate the error in terms of voltage magnitude if required, as shown in equation 8.

$$V = \frac{\%error}{100} \times V_{S1} \quad V \quad (8)$$

4.2 Summary of results

Table 2 shows the percentage error (%) in train voltage calculation, using equation 7.

5 Discussion and conclusion

5.1 Discussion

The results in table 2 suggest that either equations 1 and 3 for the equivalent traction return path impedance are satisfied when there is only one train present, as in the case of equation 1 and six trains being present, as in the case of equation 3. This does not definitively define which equivalent rail return impedance should be used. Therefore an intermediate step was introduced by assessing three trains being present. The choice for three trains is purely



Table 2: Summary of case study results.

System Voltage V DC	Type	Z_L Ω/m	$\alpha\%$	Train Current A	Substation Spacing km	Error for 1 Train and six return paths		Error for 3 Trains and six return paths		Error for 6 Trains and six return paths	
						Eqn. (1)	Eqn. (3)	Eqn. (1)	Eqn. (3)	Eqn. (1)	Eqn. (3)
						% error	% error	% error	% error	% error	% error
1500	OHL	0.0784	60	4000	2	0	1.668	-0.667	1.001	-1.668	0
750	OHL	0.1472	55	1000	1.5	0	2.705	-1.082	1.623	-2.705	0
750	3rd rail	0.0651	37	5000	3	0	17.086	-6.834	10.252	-17.086	0
750	3rd rail	0.0828	50	5000	3	0	17.112	-6.845	10.267	-17.112	0



academic and is used to define a utilisation factor of the traction return circuits, although in reality the number of trains may vary. This now shows that the error is lower when using equation 1 to derive the equivalent traction return impedance. However, whilst the error is acceptable for the overhead line railways, it is not so for the 3rd rail railway. The impedance is too low in the traction power conductor, allowing a large voltage drop to occur between the train and the electrical substation, mainly due to the high values of current drawn by the trains. Furthermore, the impedance in the traction return circuit is more dominant and the inconsistency in the assumptions of equations 1 and 3 suggest that this impedance must be modelled accurately, i.e. as per equation 2 and the split impedance method.

Therefore this type of railway should be modelled as a split impedance railway, such that the train and rail voltages are calculated correctly. Furthermore, it is noticeable that the magnitude of percentage error increases as the number of tracks occupied increases. This must also be considered when deriving the power system model, with respect to split or combined impedance method.

5.2 Conclusion

The case studies confirm that the most efficient method for solving the power system model is to combine the traction power and return conductor impedances such that the OC of the circuit is reduced to as low as reasonably possible. This will allow post process simulation of the data obtained from the real time solution and allow effective management of the computing overhead that is always the fundamental driver with computational techniques and the desired accuracy. This statement is true, with the exception of the 3rd rail traction conductor railway, where the most appropriate method would be to model the impedances of the railway separately.

Acknowledgement and disclaimer

Information and guidance given in this paper are views held by the authors. The authors would like to thank colleagues within Atkins Rail and the University of Birmingham for their support and advice in writing this paper and for their kind permission in allowing this work to be published as part of the collaborative agreement to develop a DC multi train simulator.

The authors, Atkins Rail and the University of Birmingham accept no liability to anyone for any loss or damage caused by any error or omission in the work, whether such error or omission is the result of negligence or any other cause.

References

- [1] Profillidis, V.A. (2000). *Railway Engineering*. 2nd Edition. Aldershot: Ashgate Publishing Ltd. Chapter 13 pp252. ISBN 0 – 7546 – 1279 – 1.



- [2] Goodman, CJ, (2005). *Modelling and Simulation*. The IEE Second Residential Course on REIS 2005. Wrightsons. Chapter D3. ISBN 0 – 86341 – 511 – 3.
- [3] BS EN50122-1:1998 Railway applications - Fixed installations Part 1: Protective provisions relating to electrical safety and earthing.
- [4] Kiessling, F, Puschmann, R, Smeider, A. (2001). *Contact Lines for Electric Railways*. 1st Edition. Munich Erlangen: Publicis Coproate Publishing Ltd. Chapter 1 pp33. ISBN 3 – 89578 – 152 – 5.
- [5] Bird, J. (2004). *Electrical Circuit Theory and Technology*. Revised 2nd Edition. Newnes. Chapters 30, 31, 33, 44 pp 531 – 559, 575 – 598 and 869 – 900. ISBN 0 – 7506 – 5784 – 7
- [6] Stevenson, WD Jr. (1982). *Elements of Power System Analysis*. 4th Edition. McGraw Hill. Chapters 5, pp 88 – 126. ISBN 0 – 07 – 066584 – 2
- [7] Yu, JG, (1992). *Computer Analysis of Touch Voltages and Stray Currents for DC Railways*. PhD Thesis submitted to the University of Birmingham, UK
- [8] White, RD, Zhang, Z (2003) Atkins Rail internal report. *Predictions of 50Hz induced voltage into lineside cables on 25kV railway*. Report Number: BF 5001270/RDW/001. Issue 4.0 August 2003
- [9] Case, S, (2005). *Earthing, Bonding and Stray Current Mitigation - Principles*. The IEE Second Residential Course on REIS 2005. Wrightsons. Chapter D4. ISBN 0 – 86341 – 511 – 3.
- [10] Gerald CF, Wheatley PO. (2004). *Applied Numerical Analysis*. 7th Edition. Pearson/Addison Wesley. Chapter 2 pp 88 – 100. ISBN 0 – 321 – 1909 – X



This page intentionally left blank

Catenary and autotransformer coupled optimization for 2x25kV systems planning

E. Pilo, L. Rouco & A. Fernandez

Instituto de Investigación Tecnológica,

E.T.S. de Ingeniería ICAI, Univ. P. Comillas, Madrid, Spain

Abstract

In the design process of a 2x25kV power supply, the location and sizing of autotransformers and the choice of the catenary type are usually two aspects very closely interrelated. As a result, the number of autotransformers can normally be reduced or augmented if the catenary is upgraded or downgraded respectively. In this paper, coupling equations between these two aspects are described in detail.

These equations are used to formulate a multi-attribute optimization problem in which the global efficiency of the investment is optimized. In this problem, the considered attributes are: (i) investment cost and (ii) a performance index that is defined in the paper. The optimization procedure gives a reduced set of catenary and autotransformers combinations that maximize efficiency. This optimal reduced set can be used as an input in the design process of power supply. For the solution of this problem, different dominance criteria are evaluated. Furthermore, sensitivity studies are carried out during the optimization process to help in the search of catenary and autotransformers combinations.

As a study case, the proposed optimization procedure has been used to obtain a reduced set of catenaries and autotransformers based in the catenary C-350, which has been used in the Madrid – Barcelona – French border new high-speed line.

Keywords: power supply system, high speed railways, multi-attribute optimization, long-term infrastructure planning.

1 Introduction

In the design process of a 2x25kV power supply, the location and sizing of autotransformers and the choice of the catenary type are usually two aspects very



closely interrelated. As a result, the number of autotransformers can normally be reduced or augmented if the catenary is upgraded or downgraded respectively. In this paper these relationships are analyzed and used to formulate a multi-attribute optimization problem in which the global efficiency of the investment is optimized [1].

Section 2 describes the power supply system of AC-electrified railways. In section 3 an equivalent model is used to represent bi-voltage 2x25kV and to analyze the relationships between catenary parameters and autotransformers separation. In section 4 the optimization problem is described. In section 5 the optimization procedure is applied to a real case. Finally, in section 6 the conclusions of this work are presented.

2 The power supply of AC electrified railways

2.1 General structure

Figure 1 shows the general structure of power-supply systems of AC electrified railways:

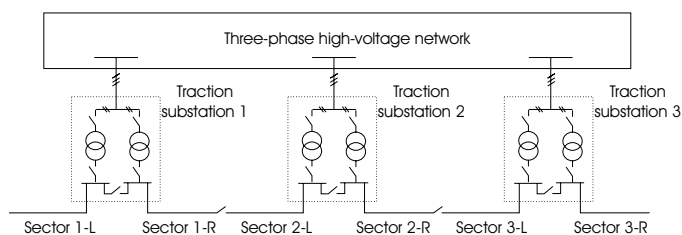


Figure 1: Structure of the power supply system.

As shown, the electrical system is divided in electrically-isolated single-phased sectors, which are fed from the three-phase network through a traction substation. These substations are connected between two of the three phases of the high-voltage network. Each of these sectors can use either mono-voltage system (1x25kV) or bi-voltage system (2x25kV). In mono-voltage systems [2], the feeding conductors are set to the specified voltage level (see Figure 2).

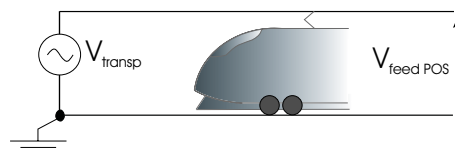


Figure 2: Mono-voltage system configuration.

In bi-voltage systems, a higher voltage is set between feeding conductors [3, 4]. This voltage is reduced by using autotransformers distributed along the



catenary (see Figure 3). In these systems, the term cell very often refers to the portion of catenary located between two consecutive autotransformers. Typical values for cell lengths are 10-15km.

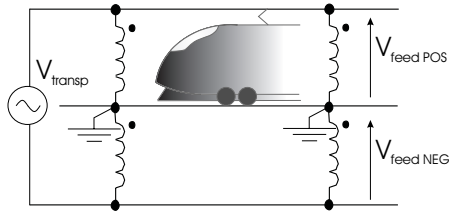


Figure 3: Bi-voltage system configuration.

As this paper is focused on this system, it is assumed that all the sectors are fed using bi-voltage system.

2.2 Catenary

The typical configuration of the catenary of an AC railway line is shown in Figure 4. The catenary contains several physical conductors that can be grouped into three groups: positive, negative and ground wires. In case of multiple tracks, other conductor arrangements are possible.

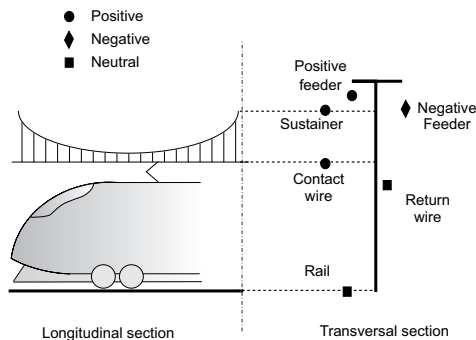


Figure 4: Typical conductor distribution.

The positive wires are the positive feeder, the sustainer wire and the contact wire. There is usually only one negative wire called negative feeder. The ground wires are the rail, the collector wire and the return wire.

The conductors of each group are connected between them at regular intervals (typically 300m). Additionally, ground conductors are frequently connected to earth.



3 The effect of autotransformers and catenary upgrades

3.1 Base magnitudes

In order to improve numerical stability, normally per unit magnitudes are used to carry out all the necessary calculations. Thus, the circuit can be divided into three zones based on their nominal voltage. Figure 5 shows the considered zones: (i) high-voltage zone, (ii) positive zone and (iii) negative zone.

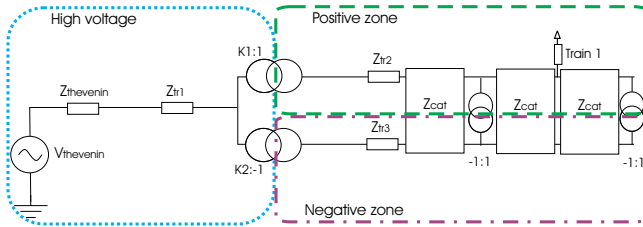


Figure 5: Zone division for base magnitudes selection.

A base power S_{base} has to be chosen and is common to all the zones (a typical value is 10MW). Furthermore, base voltages have to be selected for the three zones. If base voltages are exactly the voltages of every zone in a scenario without any kind of load, transformation ratios take values of 1 and -1 . Base impedance and base currents can be determined from the base power and voltage of each zone.

3.2 The 1x25kV equivalent model of 2x25kV systems

In [5] the behavior of bi-voltage systems is analyzed and a equivalent model is proposed to represent bi-voltage 2x25kV systems as if they were mono-voltage 1x25kV. This is the model used in the presented work.

Figure 6 shows the approximated behavior of the circuit with a train consuming a current I , assuming: (i) that voltage drop along a cell in the positive and in the negative side have the same value but different sign and (ii) that, as far as autotransformers can be supposed ideal, it can be assumed that there are current flows only in the autotransformers that are immediately adjacent to the considered train.

In this figure $V_{cell\ n}$ is the voltage drop along the cell n , $I_{p,trans}$ and $I_{n,trans}$ are respectively the positive and negative currents in the transmission cells, $I_{p,train}$ and $I_{n,train}$ are respectively the positive and negative currents in the cell of the train, L_{cell} is the length of the cell of the train, x is the relative position of the train, expressed as a fraction of L_{cell} .



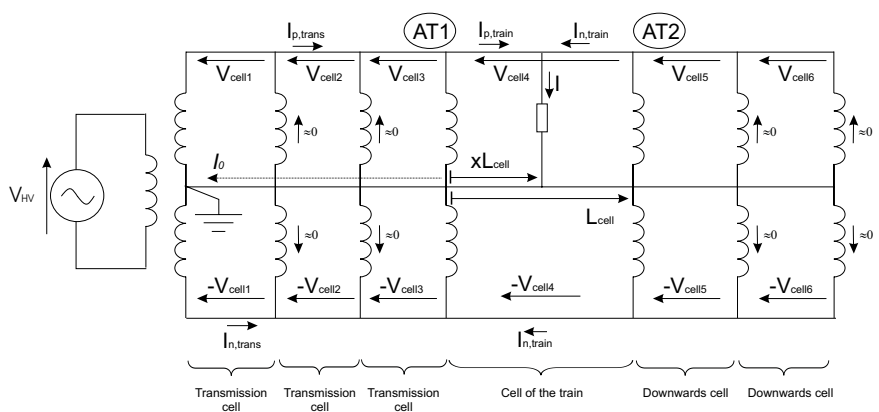


Figure 6: Approximated behavior of bi-voltage system.

Based on these simplifications, the positive phase of the bi-voltage system can be represented as shown in Figure 7. In this model, two different contributions have been identified: (i) the equivalent impedance of the catenary $z_{eqv,cat}$ that depends only on the configuration of physical conductors and (ii) z_{gap} that is associated to the separation between autotransformers.

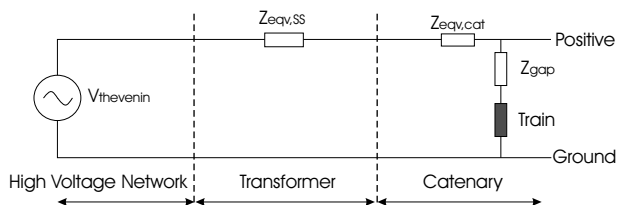


Figure 7: Mono-voltage equivalent circuit of bi-voltage system.

The parameters of this equivalent circuit are calculated as follows:

$$z_{eqv,cat} = \tilde{z}_{eqv,cat} \cdot D_{ss,train} \quad (1)$$

where the symbol $\sim$ is used to refer per length unit magnitudes and $D_{ss,train}$ is the distance between the substation and the train

The equivalent impedance $\tilde{z}_{eqv,cat}$ of the catenary can be obtained from the elements of the equivalent conductors impedance matrix [6], where the sub-indexes $_p$ and $_n$ represents the positive and negative conductors respectively.

$$\tilde{z}_{eqv,cat} = \frac{\tilde{z}_{pp} \cdot \tilde{z}_{nn} - \tilde{z}_{pn} \cdot \tilde{z}_{np}}{\tilde{z}_{pp} + \tilde{z}_{nn} + \tilde{z}_{pn} + \tilde{z}_{np}} \quad (2)$$

The impedance gap z_{gap} associated to the separation of autotransformers can be obtained as follows:



$$\mathbf{z}_{gap} = L_{cell} x(1-x) \frac{\tilde{\mathbf{z}}_{pp}^2 + \tilde{\mathbf{z}}_{pp}\tilde{\mathbf{z}}_{np} + \tilde{\mathbf{z}}_{pn}\tilde{\mathbf{z}}_{pp} + \tilde{\mathbf{z}}_{pn}\tilde{\mathbf{z}}_{np}}{\tilde{\mathbf{z}}_{pp} + \tilde{\mathbf{z}}_{pn} + \tilde{\mathbf{z}}_{np} + \tilde{\mathbf{z}}_{nn}} \tag{3}$$

The voltage drop associated to the impedance $\mathbf{z}_{gap}$ is referred as voltage deviation from the equivalent model of the catenary. As shown in Figure 8, this voltage deviation starts and ends in cell in which is located the train. In other words, at the end of this cell all the voltage deviation is recovered and thus no extra voltage drop has to be added to the trains that are located downwards.

Figure 8 summarizes the voltage drops in the sector in a scenario with only one train:

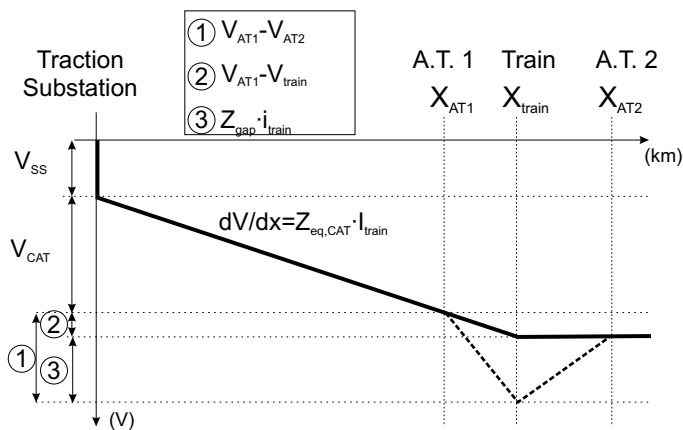


Figure 8: Voltage drops in a bi-voltage sector.

It can be seen that the deviation impedance is proportional to the distance between autotransformers. Consequently, as far as the number of autotransformers is increased, the relative weight of the deviation is reduced.

3.3 Influence of catenary type and autotransformers distance on voltages

Using the described model, sensitivity of the voltage drops to catenary upgrades or to autotransformer additions can be analyzed. These are two common investment decisions to be taken in the design process of the power supply in bi-voltage 2x25kV electrified railways.

Figure 9 shows the effect of upgrading the catenary in the voltage drops along the catenary ΔV_{cat} . When upgrading the catenary, the most important effect is a reduction of voltage drops all along the sector. Additionally, the voltage drop due the separation autotransformers (corresponding to $\mathbf{z}_{gap}$) can also be reduced.

Figure 10 shows the effect of shortening the distance between consecutive autotransformers, which is typically achieved by adding extra autotransformers to the sector. Unlike the catenary upgrade, the benefits of this enhancement are limited to the cell whose distance has been reduced, due to the local influence of voltage deviations V_{DEV} .

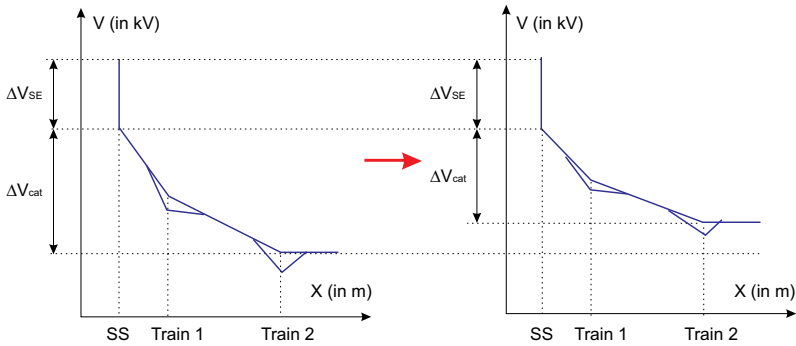


Figure 9: Effect of upgrading the catenary.

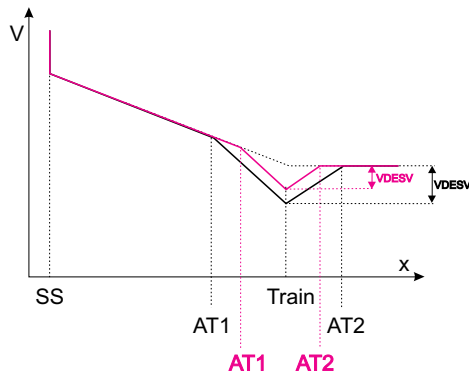


Figure 10: Effect of shortening length of cells.

As it has been described, both catenary upgrades and autotransformer additions can be used to reduce voltage drops.

4 The optimization problem

In the design process of AC-electrified railways, voltage limitations are commonly active, especially when evaluated in degraded situations (substation out of order). Thus, determining the most efficient way of reducing voltage drops is a key factor in the design of the power supply.

As it has been described, catenary upgrades and autotransformer additions are often investment decisions that are exchangeable in order to reduce voltage drops, from a technical point of view. Therefore, economical criteria have to be considered to determine the most efficient combination of catenaries and autotransformer distributions. For that reason, a multi-criteria optimization problem has been formulated and solved.

The goal of this optimization is to obtain a reduced repository of combinations of catenary types and autotransformer distributions, in which the efficiency of its elements is maximized. This set is to be used as an input in the



design of the power supply system. To determine this repository, the considered attributes are: (i) investment cost and (ii) a efficiency index that is to be defined. In order to evaluate the efficiency of each {catenary, autotransformers} combination the efficiency index **ZEQ** is defined in eqn. (4). This index corresponds to the total equivalent impedance seen between the substation output and a train located in the further cell of the corresponding sector. In order to simplify the resulting expression, it has been assumed that autotransformers are uniformly distributed.

$$\mathbf{ZEQ} = \tilde{\mathbf{z}}_{eq,cat} \left[\frac{n-1}{n} L + \frac{L}{n} x \right] + \mathbf{z}_{gap} \quad (4)$$

where n is the number of autotransformers of this sector and L is the length of this sector. Expanding $\mathbf{z}_{gap}$ in eqn. (4) becomes:

$$\mathbf{ZEQ} = \tilde{\mathbf{z}}_{eq,cat} \left[\frac{n-1}{n} L + \frac{L}{n} x \right] + (\tilde{\mathbf{z}}_{pp} - \tilde{\mathbf{z}}_{eq,cat}) x (1-x) \frac{L}{n} \quad (5)$$

As **ZEQ** depends on the relative position x of the train in its cell, one of the following criteria for can be assumed:

- The train is located in the middle of its cell ($x = 0.5$)
- The train is located where the index **ZEQ** reaches its maximum.

It should be noted that **ZEQ** incorporates the effect of both catenary type and autotransformers in an equilibrated manner, as it corresponds to their contribution in the total voltage drops.

The considered cost function is:

$$C_{cat,n} = C_{cat} \cdot L + C_{AT} \cdot n \quad (6)$$

where $C_{cat,n}$ is the cost associated to catenary *cat* with n autotransformers, C_{cat} is the per length unit cost of the catenary *cat* and C_{AT} is the unitary cost of each autotransformer.

To get the optimal repository the following steps have to be accomplished:

- Exploration of the possible catenary/autotransformer configurations. As the sustainers and the contact wires are fixed, the number of combinations is not excessive. Symmetry restrictions can also be used to reduce the search space.
- Eliminate dominated configurations (a configuration is dominated if it is worse in cost and in efficiency than another configuration). Relaxed dominance criteria can also be applied in order the eliminate configurations (i) a bit cheaper but much lower efficiency or (ii) a bit more efficient but much more expensive.
- Within the resulting repository, chose the final configurations trying to cover uniformly the costs range.



5 Study case

The proposed optimization procedure has been used to get a repository of catenaries and autotransformers based on the catenary C-350 that has been used in the Madrid – Barcelona – French border new high-speed line. The considered costs are summarized in Table 1.

Table 1: Costs structure of the study case.

Description	Cost	Observations
Autotransformer	200 units	Cost per autotransformer
Fixed cost of catenary	29 units/km	Includes installation of catenary towers, insulators, contact wires and sustainers
Cost per added feeder	0.3 units/km/feeder	Only in configurations with positive or negative feeder
LA-110 conductor	0.716 units/km	To be included only if used
LA-180 conductor	0.832 units/km	To be included only if used
LA-280 conductor	0.95 units/km	To be included only if used
LA-380 conductor	0.125 units/km	To be included only if used

Figure 11 shows the 5 catenary/autotransformers combinations chosen (marked with arrows), that have been evaluated for sector 30km, 40km and 50km long.

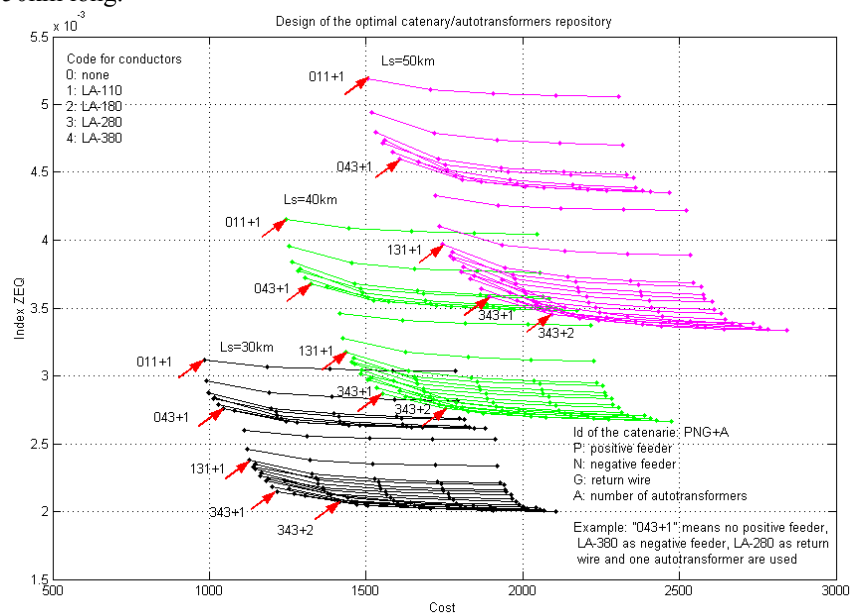


Figure 11: Effect of shortening length of cells.



Each curve represents the effect of adding up to 5 autotransformers to the catenary specified, for the given sector length.

As shown in Figure 11, the efficiency gain of upgrading catenary (change to a lower curve of the same family) is normally a more efficient option than adding an extra autotransformer (go to the following point of the same curve).

It can also be observed that the efficiency of adding autotransformers decreases very quickly and, thus, none of the chosen configurations uses more than 2 autotransformers.

6 Conclusions

Using the mono-voltage equivalent model of bi-voltage systems, the existing relationships between the location of autotransformers and the choice of the catenary have been analyzed. Furthermore, a procedure has been proposed to obtain a reduced repository of combinations of catenary types and number of autotransformers in which efficiency and cost are optimized

As a study case, the proposed optimization procedure has been used to obtain a reduced set of catenaries and autotransformers based in the catenary C-350, which has been used in the Madrid – Barcelona – French border new high-speed line. The solution obtained in the study case suggests that upgrading catenary is normally a much more efficient option than adding an extra autotransformer. The reason for that is that the marginal cost of using larger conductors in the catenary can be neglected compared to the marginal cost of adding an autotransformer.

References

- [1] E. Pilo, "Optimización del Diseño de la Electrificación de Trenes de Alta Velocidad," in *IIT*, vol. PhD: Univ. Pont. Comillas de Madrid, 2003.
- [2] J. D. Glover, A. Kusko, and S. M. Peeran, "Train voltage analysis for AC railroad electrification," *IEEE Transactions on Industry Applications*, vol. IA-20, pp. 925-934, 1984.
- [3] R. J. Hill and I. H. Cevik, "On-line simulation of voltage regulation in autotransformer-fed AC electric railroad traction networks," *IEEE Transactions on Vehicular Technology*, vol. 42, pp. 365-372, 1993.
- [4] P. H. Hsi, S. L. Chen, and R. J. Li, "Simulating on-line dynamic voltages of multiple trains under real operating conditions for AC railways," *IEEE Transactions on Power Systems*, vol. 14, pp. 452-459, 1999.
- [5] E. Pilo, L. Rouco, and A. Fernández, "A reduced representation of 2x25 kV electrical systems for high-speed railways," presented at IEEE/ASME Joint Rail Conference, Chicago, 2003.
- [6] E. Pilo, L. Rouco, A. Fernández, and A. Hernández-Velilla, "A simulation tool for the design of the electrical supply system of high-speed railway lines," presented at IEEE PES Summer meeting 2000, Seattle, 2000.



A study of capacity calculation of regenerative inverter for 1500V DC traction system

C. H. Bae, M. S. Han, Y. K. Kim, S. Y. Kwon & H. J. Park
Korea Railroad Research Institute, Korea

Abstract

The paper presents methods of determining the capacity and installation positions of regenerative inverters installed in 1500V DC electric railway system. We suggested a method that calculates using regenerative power obtained from Train Performance Simulation (TPS) and Power Flow Simulation (PFS). We carried out TPS and PFS for Seoul Subway line 5 and 6, calculating regenerative power and determining substations where regenerative inverters would be installed and the optimal capacity and number of inverters to be installed.

Keywords: regenerative inverter, electric railway system, train performance simulation, power flow simulation.

1 Introduction

In DC electric railway system, 22.9kV system voltage is converted to DC 1500V voltage through a 3-phase silicon diode rectifier and supplied to motor cars. Because, in case of diode rectifiers, regenerative power generated at the regenerative braking of motor cars cannot be absorbed into the supply grid, the power is used at nearby powering trains or consumed as heat at resistances mounted on the cars. However, if a regenerative inverter is installed in inverse-parallel with the diode rectifier, it absorbs dump regenerative power consumed at the resistance and transmits it to a high-voltage switchboard for reuse. In this way, energy can be saved by reusing dump regenerative power wasted away as heat and the braking and ATO performance of motor cars can be improved through enhancing the regenerative power absorption rate of catenary lines. Despite these advantages, regenerative inverters cannot be installed in all substations for electric railways because the cost of manufacturing and installing regenerative inverters is higher than the benefit from the use of regenerative



powers. Thus, they are installed at sections with a long continuous slope or where regenerative power loss in resistor bank becomes a problem, and for this it is necessary to calculate accurate regenerative power generated and determine the appropriate installation positions, number and capacity of regenerative inverters.

The present study suggests methods of determining the capacity and installation positions of regenerative inverters installed in 1500V DC electric railway system. We suggested a method that approximates using parameters related to substations where regenerative inverters are installed, railway lines and operating motor cars, and another that calculates using regenerative power obtained from Train Performance Simulation (TPS) and Power Flow Simulation (PFS). For TPS and PFS, we used a program developed by Korea Railroad Research Institute for the development of light rail transit system [1]. We carried out TPS and PFS for Seoul Subway Line 5 and 6, calculating regenerative power and determining substations where regenerative inverters would be installed and the optimal capacity and number of inverters to be installed.

2 Calculating the capacity of regenerative inverters

A regenerative inverter detects the rise of catenary line voltage caused by dump regenerative power, absorbs the regenerative power, and transmits it to a high-voltage distribution system for reuse. Fig. 1 shows a diode rectifier and a regenerative inverter at an electric railway substation. Here, the 12-pulse diode rectifier generates 1500V DC voltage and the IGBT regenerative inverter absorbs dump regenerative power and transmits it to the supply grid for reuse. As for the capacity of regenerative inverters, because if the operating interval of motor cars is very long the instantaneous load rate is even higher than that of a rectifier supplying DC 1500V power, we need a very high peak power rating. However, high peak power rating brings restrictions on the economic efficiency and size of regenerative inverters. Accordingly, maximum power higher than a certain level effectuates regenerative power by allowing current limiting and DC voltage rise to regenerative inverters, and acknowledges it as peak power rating.

To estimate the optimal capacity of such a regenerative inverter, it is desirable to block regenerative power loss in an operating train, make a route for absorbing regenerative power and measure regenerative power, but because this requires additional regenerative power absorbing equipment and operation to prevent regenerative power loss, it is not easy to measure regenerative power using this method. There are other methods such as approximating based on variables related to substation, operating line, train condition and regenerative power in other lines and calculating using TPS and PFS. However, because the level of regenerative power varies according to the conditions of line on which the regenerative inverter is installed, train condition and operation condition, it is difficult to determine the accurate capacity through approximation based on these major variables. Accordingly, we need to calculate dump regenerative power in various train operation conditions by conducting TPS and PFS under different conditions of line, train and substation.



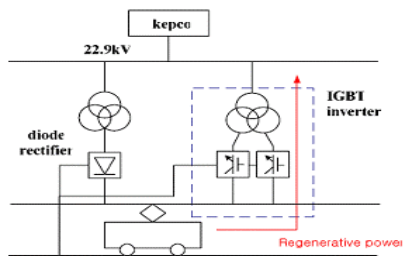


Figure 1: DC 1500V traction substation with a regenerative inverter.

3 Approximation method

Fig. 1 shows the layout of a substation for DC electric railway for calculating the capacity of a regenerative inverter, and conditions for the calculation are presented in table 1. As in fig. 1, a regenerative inverter in charge of a 12km-long regeneration section is installed at substation B, and the number of trains running in the section, n , is obtained by eqn. (1).

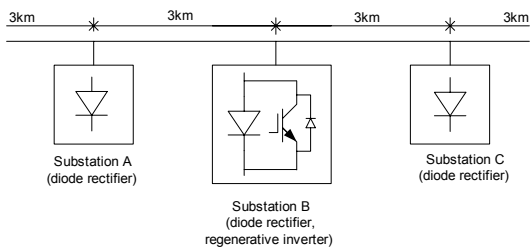


Figure 2: DC 1500V traction power system.

$$n = \frac{l}{v_s \cdot t_h} \quad [\text{trains/h}] \tag{1}$$

Maximum power consumption per hour, P_m , is calculated from train ton-kilo capacity as follows.

$$P_m = 2n \cdot s \cdot w \cdot l \cdot (1 + a) \cdot k \quad [\text{kW}] \tag{2}$$

Here, coefficient 2 means a double track section, and a is standard deviation of power variation according to DIA distribution. Capacity is calculated using power regeneration rate and regenerative braking efficiency rate obtained from existing substations with a regenerative inverter. Power regeneration rate, λ_1 ,

means the ratio of absorbed regenerative power to P_m , the maximum power consumption of substations with a regenerative inverter, and regenerative braking efficiency rate means the ratio of absorbed regenerative power to the whole regenerative power generated within the section covered by a substation with a regenerative inverter. Here, the whole regenerative power includes regenerative power consumed by nearby powering trains and regenerative power loss in resistor bank. In general, power regeneration rate λ_1 range between 0.20 ~ 0.23, and regenerative braking efficiency rate λ_2 between 0.63 ~ 0.67 [2]. Using these data, the capacity of a regenerative inverter can be calculated by eqn. (3), where W is the whole regenerative power generated from the section covered by the regenerative inverter, which includes regenerative power consumed by nearby powering trains and regenerative power loss in resistor bank. Accordingly, the capacity of the regenerative inverter should be larger than W considering the operation condition of the line.

$$W = P_m \times \frac{\lambda_1}{\lambda_2} \text{ [kW]} \quad (3)$$

In order to calculate regenerative peak current I_b from mechanical energy absorbed at braking, consumed braking power P_b is calculated by eqn. (4), and braking electric power generated from a train at speed of v [km/h] is calculated by eqn. (5).

$$F_b = (31 \times \beta - r) \times s \times w \text{ [kgf]} \quad (4)$$

$$P_b = \frac{F_b \times v}{367} \times \eta \text{ [kW]} \quad (5)$$

Here, η is power delivery efficiency. Consumed braking power P_b is obtained from eqn. (4) and (5), and regenerative peak current can be calculated as follows.

$$I_b = \frac{P_b}{V_{inv}} \text{ [kA]} \quad (6)$$

In order to estimate the optimal capacity of a regenerative inverter, we calculated W regenerative power generable from the covered section, and I_b regenerative peak current. From the conditions in Table 1 were obtained $W=1480$ [kW] and $I_b=3.5$ [kA], and the capacity of a regenerative inverter approximated 1.5MVA, 350% 1 minute. Because the capacity of a regenerative inverter determined through the approximation method does not consider line conditions and train operation conditions, it can be used only to review the whole system capacity rather than as a specification to install a regenerative inverter.



Table 1: Calculation conditions.

Item	Value	Item	Value
Number of cars, s	8 (4M4T)	Running resistance, r	10kg/ton
Headway, t_h	2.5 min	Maximum speed, v_m	80km/h
Weight, w	48ton/car	Commercial speed, v_s	35km/h
Decelerating speed, β	3.5km/h/s	Regenerative operation voltage, V_{inv}	1650V
Train ton-kilo capacity, k	50kW/1000ton.km	Power regeneration rate, λ_1	0.20
Power delivery efficiency, η	0.85	Regenerative braking efficiency rate, λ_2	0.65

4 PFS method

This section explains how to determine the capacity of a regenerative inverter using TPS and PFS. After PFS is performed with changing the capacity and the number of regenerative inverters to be installed, the regenerative power loss of the line is calculated. Here, the loss ratio of regenerative power means the ratio of regenerative power consumed as heat in the cars to the whole regenerative power generated as shown in eqn. (7). After the optimal position and the number of regenerative inverters are determined in a way of reducing the calculated loss ratio of regenerative power to the maximum, the root mean square of regenerative power (RMS power) and peak power are calculated. The effective regenerative power per hour calculated by eqn. (8) determines the continuous rating of regenerative inverter, and is used to determine peak power rating based on the maximum regenerative power rate and the braking time of motor cars.

regenerative power loss

$$= \frac{\text{regenerative power} - \text{output power of regenerative inverter}}{\text{regenerative power}} \tag{7}$$

$$P = \sqrt{\frac{1}{T_s} \int_{t_1}^{t_2} p_{reg}(t)^2 dt} \tag{8}$$

Here, P is the root mean square of regenerative power $P_{reg}(t)$, and T_s was 1 hour. The method of determining the position, number and capacity of regenerative inverters to be installed is shown in the block diagram in fig. 3, and its details are as follows.

1. Perform PFS for the case that regenerative inverters are installed in all substations on the line.
2. Calculate the mean square of regenerative power of each substation, and rank the substations according to regenerative power.

3. Perform PFS after removing regenerative inverters from the two substations with the lowest regenerative power.
4. Again calculate the root mean square of regenerative power of each station with a regenerative inverter, and calculate the loss ratio of regenerative power for the whole line.
5. Perform PFS while removing regenerative inverters one by one from substations with the lowest regenerative power.
6. Draw the curve of the loss ratio of regenerative power according to the number of regenerative inverters installed in substations, and selects the curve that shows the largest reduction in regenerative power loss.

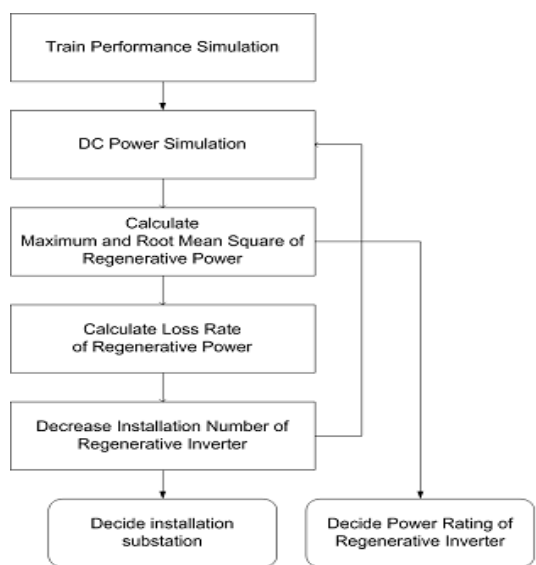


Figure 3: Flowchart for substation selection.

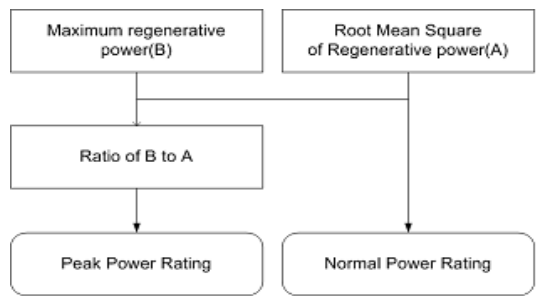


Figure 4: Flowchart for regenerative inverter capacity.

Once the position and number of regenerative inverters to be installed are determined, the rated capacity of regenerative inverter and the peak power

capacity are calculated through the procedure in fig. 4. The rated capacity of a regenerative inverter sets the root mean square value of regenerative power obtained from substations, and peak power rating is determined by the ratio of the peak regenerative power to the root mean square value of regenerative power. In addition, because time for the rise of catenary line voltage caused by dump regenerative power of subway substations does not exceed 1 minute, peak power rating is assumed to continue for 1 minute.

We performed TPS and PFS using data on trains and lines of Seoul Subway Line 5 and 6. Figs. 5 and 6 show the catenary line voltage and the power consumption waveform of substations according to whether a regenerative inverter is installed or not. In the fig. 5, regenerative power generated by the power braking of motor cars is raising catenary line voltage instantaneously. Fig. 6 shows that regenerative power is absorbed by the substation and the variation of catenary line voltage is reduced.

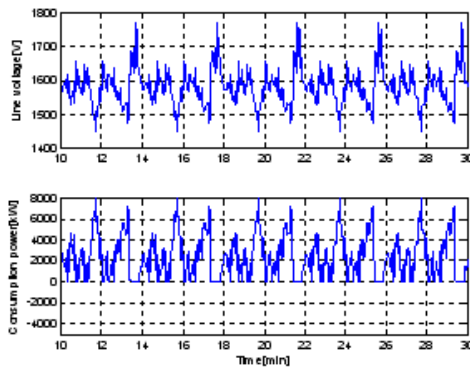


Figure 5: Seoul Line 6 Substation 8 without a regenerative inverter.

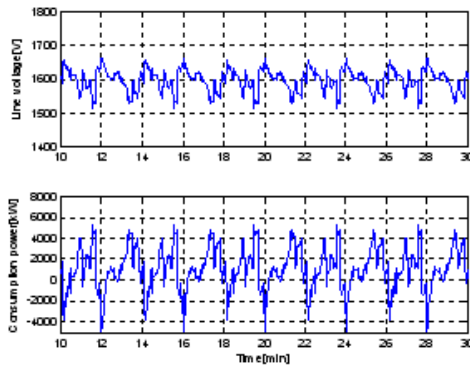


Figure 6: Seoul Line 6 Substation 12 with a regenerative inverter.



Fig. 7 shows absorbed regenerative power according to the number of substations with a regenerative inverter. Fig. 7 (a) is the case that regenerative inverters are installed in all substations. Regenerative power is different among substations because of differences in the gradient of line, distance between stations and training operation conditions. Figs. 7(b)~(f) show regenerative power of each substation while removing regenerative inverters one by one from the substation with the lowest regenerative power. As the number of substations with a regenerative inverter decreases, regenerative power at nearby substations with a regenerative inverter increases to some degree.

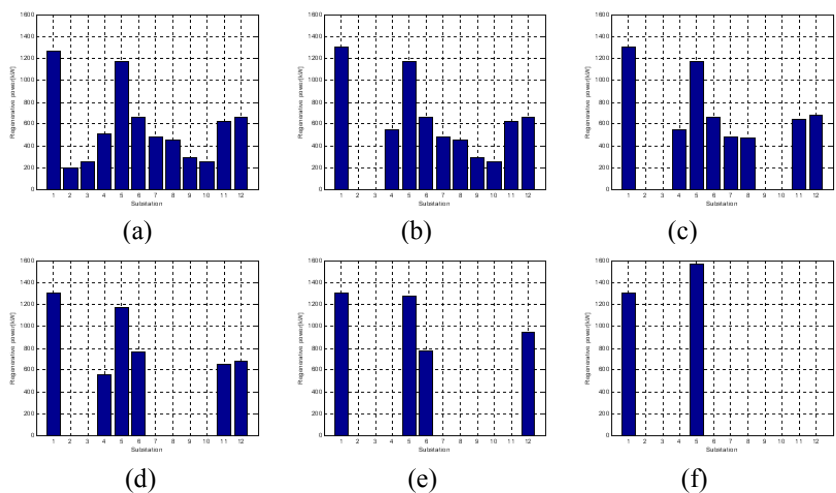


Figure 7: RMS of regenerative power in Seoul Line 6.

Because the unit price of a regenerative inverter and the initial cost of installation are high, it is economically inefficient to install regenerative inverters at all substations. Accordingly, we should find the position and number of substations to install regenerative inverters centering on substations with high regenerative power as shown in fig. 4, in a way of reducing the regenerative power loss of the whole line to the maximum.

Figs. 8 and 9 shows the curve of loss ratio of regenerative power changing according to the number of regenerative inverters in Seoul line 5 and 6. As a large-capacity regenerative inverter makes it possible to transmit more regenerative power to the supply grid, the loss ratio of regenerative power is reduced, and the curve of regenerative power loss goes down with the increase in the number of regenerative inverters installed. However, the reduction rate of regenerative power loss is not constant. It is because regenerative power is different among substations. As shown in figs. 8 and 9, reduction in the loss ratio of regenerative power decreases gradually with the increase in the number of substations with a regenerative inverter.

In case of Seoul Line 6, reduction in the loss ratio of regenerative power is largest when regenerative inverters installed at four substations. Because larger reduction in regenerative power loss is not expected from the installation of more regenerative inverters, it is desirable to install four regenerative inverters. As in fig. 7, the adequate capacity of regenerative inverter for Substation 1 and 5 is 1.5MVA and 1MVA for Substation 6 and 12. However, because Substation 5 and 6 are neighboring to each other, it is economically more efficient to install a regenerative inverter only at Substation 5 than at both.

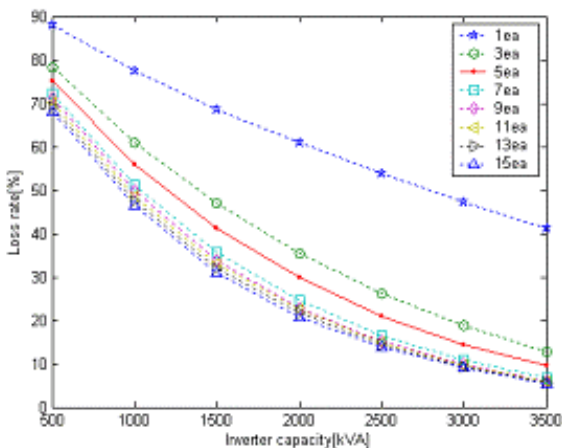


Figure 8: Loss rate of regenerative power in Seoul Line 5.

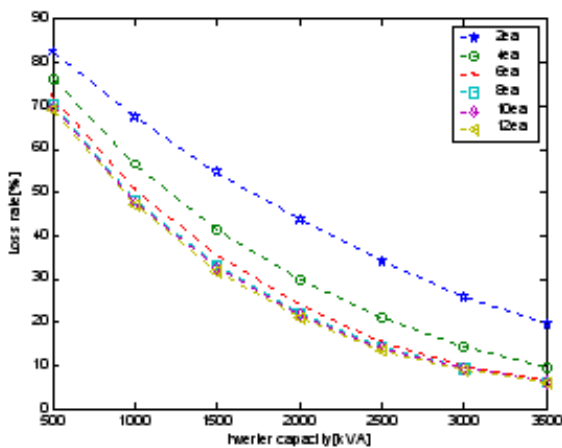


Figure 9: Loss rate of regenerative power in Seoul Line 6.



We performed PFS for Seoul Subway line 5 and 6, and presented the results in table 2. The optimal capacity of regenerative inverter is determined by estimating rated capacity larger than the root mean square of regenerative power from each substation and determining peak power rating using the ratio of peak regenerative power to the rated capacity.

Table 2: Power simulation results of Seoul subways.

Line	Substation	RMS of regenerative power[kW]	Peak regenerative power[kW]	Ratio [%]
5	Euljiro 4-ga	1449	7102	490
	Haengdang	1284	5664	441
	Majang	1350	6554	485
6	Eungam	1305	6780	520
	Daeheung	1279	6481	507
	Samgakji	780	3827	491
	Shinnae	941	4833	514

5 Conclusions

The present study suggested a method of determining the position, number and optimal capacity of regenerative inverters that must be considered in installing regenerative inverters for DC 1500V electric railway system. We determined the capacity of regenerative inverter through simple calculation using the conditions of substations and train operation and the regeneration rate of other railway lines. We calculated the loss ratio of regenerative power and the root mean square of regenerative power for each substation by performing TPS and PFS, and determined the position and number of regenerative inverters considering change in the loss ratio of regenerative power. Applying TPS and PFS to Seoul Subway line 5 and 6, we obtained the optimal position, number and capacity of regenerative inverters to be installed.

References

[1] S.K. Jung et al., “Right Rail transit system development”, Korea Railroad Research Institute, 2002.

[2] Electric Railway DC Power Supply System Investigation Committee, “Phenomena of Power Supply System Including Regenerative Cars and Future Directions”, Technological Report No. 296 of Japanese Institute of Electrical Engineers, 1989



Train operation minimizing energy consumption in DC electric railway with on-board energy storage device

K. Matsuda, H. Ko & M. Miyatake
Sophia University, Japan

Abstract

The optimal train operation which minimizes sum of supplied energy from substations is presented in this paper. In recent years, the energy storage devices have enough energy and power density to use in trains as on-board energy storage. The electric double layer capacitor (EDLC) is assumed as an energy storage device in our study, because of its high power density. The on-board storage can assist the acceleration/deceleration of the train and may decrease energy consumption. Many works on the application of the energy storage devices to trains were reported, however, they did not deal enough with the optimality of the control of the devices. On the other hand, our previous works were to optimize acceleration/deceleration commands of the train for minimizing energy consumption without the energy storage device. Therefore, we intend to optimize acceleration/deceleration commands together with current commands through energy storage devices as our next research target. The proposed method can determine the optimal acceleration/deceleration and current commands at every sampling point. For this purpose, the optimal control problem of the train operation is formulated mathematically. It is generally difficult to solve the problem because the problem is composed of a large-scale non-linear system. However, the Sequential Quadratic Programming (SQP) can be applied to solve the problem. Two results with and without on-board energy storage device are compared. These optimized results indicate that the total energy consumption is reduced by at least 0.35% by using the EDLC. The relation between internal resistance and energy consumption is also revealed.

Keywords: electric double layer capacitor (EDLC), optimal control, energy saving operation, SQP method.



1 Introduction

In recent years, the energy storage devices have enough energy and power density to use in trains as on-board energy storage. The devices are for instance, a secondary battery and an Electric Double Layer Capacitor (EDLC). Above all, the EDLC has advantages such as maintenance free, long lifetime, rapid charge/discharge with large current and high efficiency. Therefore, the EDLC is the most suitable to equip trains as an auxiliary power supply. The on-board EDLC is useful because of the following two reasons. Firstly, it decreases the loss of circuit resistance by compensating voltage drop. Secondly, it enables us to utilize and recycle regenerative power efficiently and prevent regenerative failure.

Many works on the application of the energy storage devices to trains were reported. However, from an energy-saving point of view, they did not deal enough with the optimality of the control of the devices. On the other hand, our previous works [1,2] was to optimize notch commands which determine the acceleration/deceleration force in the train without energy storage devices. We optimize notch commands together with charge/discharge commands with making use of the experience of our previous study. It is significant to investigate the optimal charge/discharge command minimizing energy consumption in order to maximize the effect of installing the EDLC.

In this paper, we intend to formulate the optimal control problem of the train operation to find notch and charge/discharge commands which minimize amount of consumed energy, propose how to solve it, discuss the optimized results and find knowledge of the optimal operation. The knowledge will be applied to the future charge/discharge controllers for EDLCs.

2 Modeling of DC feeding Circuit

We modeled a DC feeding circuit when there is only one train between substations. The model circuit appears in fig. 1. In this figure, V_s and R_0 are the supply voltage and the internal resistance at a substation respectively. The values of R_1 and R_2 are wire resistances. These resistance values are proportional to the distance between the train and substation position. Positions of substations and stations are shown in fig. 2. The constants C and R_c are the capacitance and internal resistance in the capacitor respectively.

It is necessary to convert voltage by using a chopper because the voltage difference is high between the pantograph and capacitor. The chopper characteristic is too complicated to be examined in detail here. Therefore, we solved the circuit equation on the assumption that the chopper efficiency is 95%.

In addition, the energy consumption in the train is regarded as constant in short time because acceleration/deceleration commands do not change often. The motor-inverters of the train are modeled as a current load that helps solving circuit equations simply.



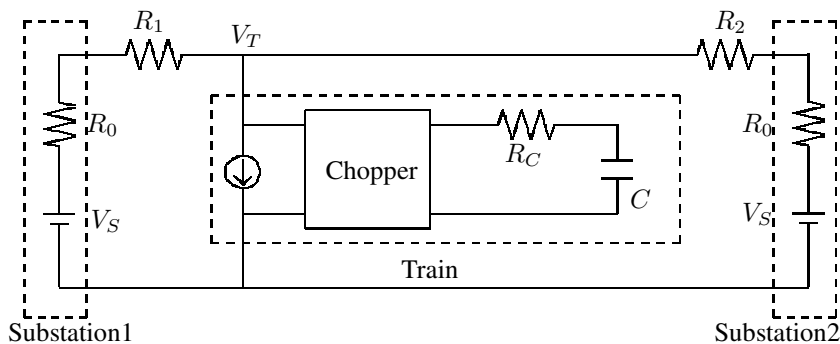


Figure 1: Circuit model with one train between substations.

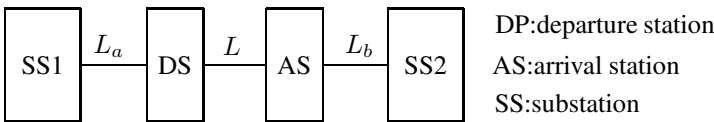


Figure 2: Positions of stations and substations.

3 Formulation of optimal control problem

We formulated the optimal control problem in this section. Here, variables are defined as follows. Control inputs n and u determine the acceleration/deceleration force and charge/discharge current through the capacitor, respectively. They are defined as table 1. State variables x , v and V_c indicate the train position, speed and capacitor voltage, respectively. Variable V_T is the voltage at the pantograph. In fact, it is a state variable if control inputs are determined and the circuit equation can be solved. However, we defined V_T as the auxiliary variable because it is difficult to solve circuit equations analytically. Additionally, these all variables depend on time t . The optimal control problem is described as follows, mathematically.

Table 1: Definition of control inputs n and u .

n or u	Operation mode	Current through the capacitor
-1	maximum deceleration	maximum charge
negative	deceleration	charge
0	coast	wait
positive	acceleration	discharge
1	maximum acceleration	maximum discharge



Minimizing the objective function

$$J = \int_0^T V_s I_s(x, V_T) dt \quad (1)$$

Subject to the following equality and inequality constraints

$$\dot{x} = v \quad (2)$$

$$\dot{v} = f(n, v, V_T) - r(v) \quad (3)$$

$$\dot{V}_c = -I_c(u)/C \quad (4)$$

$$P_T(n, v, V_T) = P_s(x, V_T) + P_c(u, V_c) \quad (5)$$

$$x(0) = 0 \quad v(0) = 0 \quad V_c(0) = V_{c\_first} \quad (6)$$

$$x(T) = L \quad v(T) = 0 \quad V_c(T) = V_{c\_final} \quad (7)$$

$$-1 \leq n \leq 1 \quad (8)$$

$$-1 \leq u \leq 1 \quad (9)$$

$$V_{T\_min} \leq V_T \leq V_{T\_max} \quad (10)$$

$$V_{c\_min} \leq V_c \leq V_{c\_max} \quad (11)$$

$$0 \leq x \leq L \quad (12)$$

$$v \geq 0 \quad (13)$$

where

I_s	sum of load currents supplied by substations
I_c	current through the capacitor
f, r	acceleration/deceleration force and running resistance per kg
P_T	electric power supplied to motor-inverters of the train
P_s, P_c	power from substations and the capacitor
V_{T_min}, V_{T_max}	lower and upper limitations of the voltage at the pantograph
V_{c_min}, V_{c_max}	lower and upper limitations of the capacitor voltage
V_{c_first}, V_{c_final}	first and final values of the capacitor voltage
L, T	distance and running time between the departure and arrival station.

The objective function is sum of supplied energy by two substations given as eqn. (1). Equality constraints are given as eqns. (2-7). Eqns. (2),(3) are motion equations of the train. The capacitor voltage is given as the eqn. (4). As mentioned above, we must solve the circuit eqn. (5) because we defined V_T as an auxiliary variable. Eqns. (6),(7) describe the initial and final conditions of state variables. Inequality constraints of control inputs, state and auxiliary variables are shown in eqns. (8-13). Especially, we did not consider speed limitations in eqn. (13).



We defined functions as below.

$$P_T(n, v, V_T) = \begin{cases} Mvf(n, v, V_T)m_e & (n \geq 0) \\ Mvf(n, v, V_T)/g_e & (n \leq 0) \end{cases} \quad (14)$$

$$P_s(x, V_T) = \left(\frac{V_s - V_T}{R_0 + R_1(x)} + \frac{V_s - V_T}{R_0 + R_2(x)} \right) \quad (15)$$

$$R_1(x) = (L_a + x)r_0 \quad R_2(x) = (L - x + L_b)r_0 \quad (16)$$

$$P_c(u, V_c) = \begin{cases} V_c I_c(u) c_e & (u \geq 0) \\ V_c I_c(u) / c_e & (u \leq 0) \end{cases} \quad (17)$$

$$I_c(u) = u I_{c\_max} \quad (18)$$

Here, m_e and g_e are motor/generator efficiency. Wire resistances R_1 and R_2 are given in eqn. (16) when the position of the departure station is defined as $x = 0$. The constants L_a and L_b indicate the distance from the departure and arrival station to the substation1 and substation2 shown in fig. 1. The constant r_0 is the wire resistance per meter. The constant c_e is the chopper efficiency. The constant I_{c_max} is the rated value of the current from the capacitor.

Additionally, maximum acceleration/deceleration characteristics, such as the control input n is 1 or -1, and running resistance are given in fig. 3. these characteristics are influenced by the voltage at the pantograph V_T . Especially, we assume that the braking system is the air supplement control. In short, the use of electrical and mechanical blended braking system is considered if the regenerative braking force is not enough for the specific braking force. Moreover, we are not concerned with the characteristic of the squeezing control because we also assume the regenerative power can be absorbed at substations.

The absorbed power can be accounted in eqn. (1).

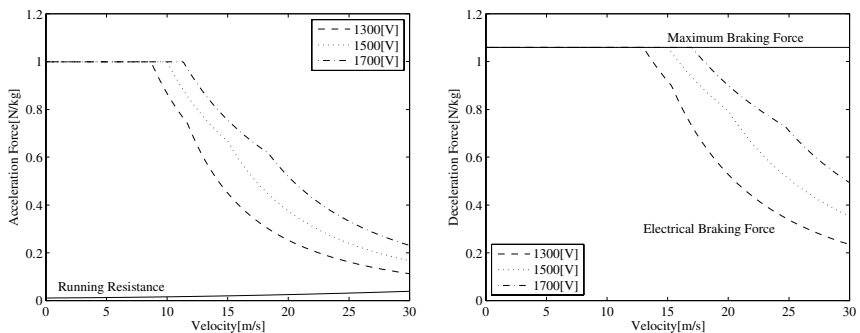


Figure 3: Acceleration/deceleration characteristics and running resistance.



4 Optimization method

We show the optimization method for solving the optimal control problem. The optimal control problem is solved by Sequential Quadratic Programming (SQP). SQP is an optimization method to solve a general nonlinear programming problem. A general optimal control problem with equality and inequality constraints can be written as

$$\begin{aligned} \text{Minimize :} \quad & J(\boldsymbol{\omega}) \\ \text{Subject to :} \quad & g_i(\boldsymbol{\omega}) = 0 \quad (i = 1, \dots, n_g) \\ & h_j(\boldsymbol{\omega}) \leq 0 \quad (j = 1, \dots, n_h) \end{aligned} \quad (19)$$

where $\boldsymbol{\omega} = (\mathbf{n}, \mathbf{u}, \mathbf{V}_T, \mathbf{x}, \mathbf{v}, \mathbf{V}_c)^\top$ is the vector of variables, J is the objective function, g and h are equality and inequality constraints, n_g and n_h are the number of equality and inequality constraints.

Next, the objective function is expressed using the second order approximation around the feasible point $\boldsymbol{\omega}^{(k)}$. Similarly, equality and inequality constraints are also expressed as the first order approximation around the feasible point $\boldsymbol{\omega}^{(k)}$ in problem (19). The transformed problem is shown as

$$\begin{aligned} \text{Minimize :} \quad & \frac{1}{2} \mathbf{d}_k^\top \mathbf{B}_k \mathbf{d}_k + \nabla J(\boldsymbol{\omega}^{(k)}) \mathbf{d}_k \\ \text{Subject to :} \quad & g_i(\boldsymbol{\omega}^{(k)}) + \nabla g_i(\boldsymbol{\omega}^{(k)}) \mathbf{d}_k = 0 \quad (i = 1, \dots, n_g) \\ & h_j(\boldsymbol{\omega}^{(k)}) + \nabla h_j(\boldsymbol{\omega}^{(k)}) \mathbf{d}_k \leq 0 \quad (j = 1, \dots, n_h) \end{aligned} \quad (20)$$

where $\mathbf{d}_k = \boldsymbol{\omega} - \boldsymbol{\omega}^{(k)}$, $\mathbf{B}_k$ is positive definite matrix. In general, the problem (20) can be solved by the interior point method [3, 4]. The optimization result of the problem (20) is a search direction $\mathbf{d}_k$.

Here, we define merit function as

$$\psi(\boldsymbol{\omega}) = J(\boldsymbol{\omega}) + \mu \left[\sum_{i=1}^{n_g} |g_i(\boldsymbol{\omega})| + \sum_{j=1}^{n_h} \max(h_j(\boldsymbol{\omega}), 0) \right] \quad (21)$$

where μ is a large positive constant. Finally, we have to find α to minimize $\psi(\boldsymbol{\omega}^{(k+1)})$ where $\boldsymbol{\omega}^{(k+1)} = \boldsymbol{\omega}^{(k)} + \alpha \mathbf{d}_k$. The vector $\boldsymbol{\omega}^{(k+1)}$ is the next feasible point. Consequently, The optimization problem (19) can be solved by iterating the following procedure.

Step 1) give the initial feasible point $\boldsymbol{\omega}^{(0)}$, and set $k = 0$

Step 2) solve problem (20), and obtain a search direction $\mathbf{d}_k$

Step 3) stop iteration if the norm $\|\mathbf{d}_k\|$ is less than 10^{-6} .

Step 4) find α minimizing the merit function, and obtain the next feasible point $\boldsymbol{\omega}^{(k+1)}$

Step 5) set $k = k + 1$, and return step 2



5 Optimization result

Optimization results are presented in this section. Specific parameters are presented as table 2. Here, the continuous time formulation must be transformed to a discrete time program in order to apply SQP to the optimal control problem. Therefore, Δt is defined as the sampling interval. In addition, the final capacitor voltage is given to equal the initial one. So, the capacitor does not have to be charged when the train arrives at the station.

We show three optimization results when the train run on straight line without speed limitations and gradients. Case 1 is the optimization result of the train without the capacitor. Case 2 and Case 3 show results of the sensitive analysis in case of the constant R_c is $0.3[\Omega]$ or $0.03[\Omega]$ respectively. These three optimization results are shown in table 3.

Case 1: The optimization result is shown in fig. 4. The optimal train operation consists of the maximum acceleration, reduced acceleration by degrees, coasting and maximum deceleration. In previous works [1, 2], we obtained the similar optimization result. Therefore, these results indicate the reliability of the proposed method.

Case 2: fig. 5(a) shows the optimization result in case that R_c is $0.3[\Omega]$. The optimal control input n do not differ from one in case 1. Next, we examine charge/discharge characteristics of the capacitor. As supplied power to the train is higher, discharge current is higher. Similarly, this pattern of the charge characteristic is also represented when the train decelerates. Qualitatively, this

Table 2: Specific parameters.

Operating condition		capacitor parameters	
$T, \Delta t$	130[s], 1[s]	C	32.3[F]
L	2000[m]	$V_{c,max}$	560[V]
R_0	$0.03[\Omega]$	$V_{c,min}$	300[V]
V_s	1500[V]	$V_{c,first}$	560[V]
r_0	$0.04 \times 10^{-3}[\Omega/m]$	$V_{c,final}$	560[V]
L_a, L_b	5000[m]	$I_{c,max}$	500[A]
M	$250 \times 10^3[\text{kg}]$	R_c	$0.3[\Omega]$ or $0.03[\Omega]$

Table 3: Optimization results.

	Total energy consumption [MJ]	Energy-saving effect[%]
Case 1	27.55	-
Case 2	27.45	0.35%
Case 3	26.74	2.92%



characteristic is proper, because it is the most effective control to compensate the voltage drop at the pantograph and prevent the regenerative failure. In Case 2, despite the lower limit value of the capacitor voltage $V_{c,min}$ set to 300[V], the capacitor stops discharging when the value of the capacitor voltage drops to about 480[V]. This result is attributed to the higher interior resistance of the capacitor. As a result, the efficiency of the capacitor itself is severely down according to the voltage drop of the capacitor.

Case 3: The optimization result is shown as fig. 5(b). In this case, there is also little variation in control input n . However, the optimal charge/discharge command includes two significant difference compared to Case 2. Firstly, the capacitor voltage reaches the value of the lower limitation when control input n changes from acceleration to coast at about time 50[s]. Secondly, substations supply the power with a small current for charging the capacitor when the train coasts. It is found from the result that the capacitor is utilized effectively. Finally, the total energy consumption is 26.74[MJ]. Compared with Case 1, the total energy consumption is reduced about 2.92%. The result indicates that energy-saving effect is higher if the interior resistance of the capacitor is lower in the future.

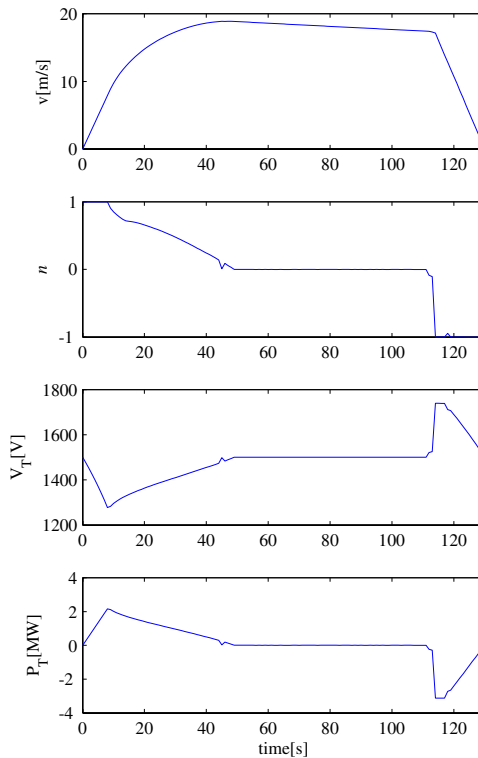


Figure 4: Optimization result (case 1).



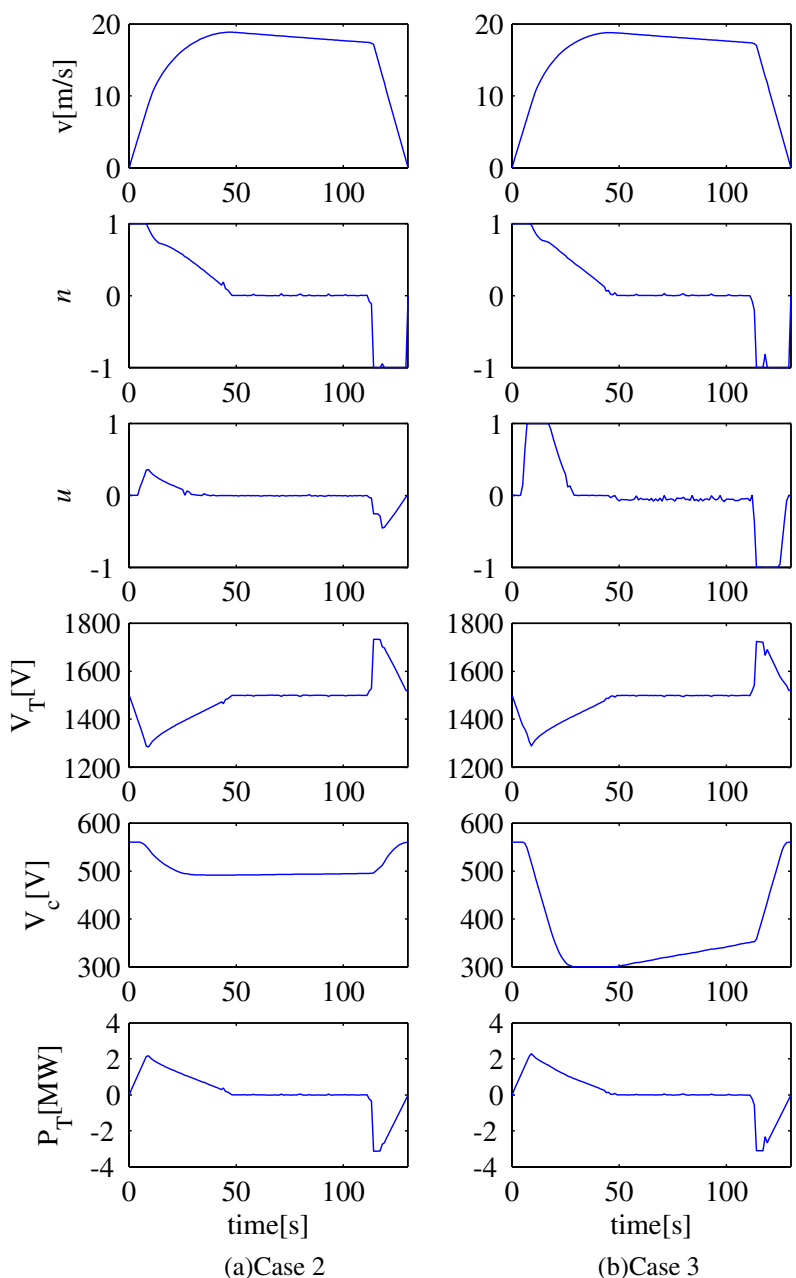


Figure 5: Optimization results.



6 Conclusion

This paper presents the optimal train operation with EDLC minimizing energy consumption. SQP can be applied to the formulated optimal control problem with discrete-time transformation. As a result, it is found that the energy consumption supplied from substations can be reduced by using on-board energy storage device effectively. Compared with the train without EDLC, the total energy consumption is reduced by 0.35% and 2.92% in Case 1 and Case 2 respectively. It is also clarified that the EDLC should not fully discharge when the internal resistance is high.

A further direction of this study will be to optimize the train operation problem which has more complicated running conditions, for example, speed limitations and gradients.

References

- [1] H. Ko, T. Koseki and M. Miyatake: "Numerical Study on Dynamic Programming Applied to Optimization of Running Profile of a Train", IEEJ Transactions on Industry Applications, Vol.125-D, No.12, pp. 1084-1092, 2005 (in Japanese).
- [2] H. Ko, T. Koseki and M. Miyatake, "Application of Dynamic Programming to Optimization of Running Profile of A Train", Computers in Railways, WIT Press, pp. 103-112, 2004.
- [3] Imad M. Nejdawi, Kevin A. Clements and Paul W. Davis, "an efficient interior point method for sequential quadratic programming based optimal power flow", IEEE Trans. on Power Systems, vol.15, NO.4, November 2000, pp. 1179-1183.
- [4] G. Irisarri, L. M. Kimball, K. A. Clements, A. Bagchi and P. W. Davis, "Economic Dispatch with Network and Ramping Constraints via Interior Point Methods" IEEE Trans. on Power Systems, Vol.13, No.1, February 1998



Calculations and measurements of harmonic current distributions in the catenary of railways with single-phase A.C.

A. Zynovchenko¹, J. Xie¹, S. Jank² & F. Klier²

¹*Department of Energy Conversion and Storage, University of Ulm, Germany*

²*Department of Electromagnetic Compatibility, DB Systemtechnik, Munich, Germany*

Abstract

Frequency converters find wide application in rail vehicles and power supply stations of railway networks. An important disadvantage of frequency converters is the emission of voltage and current harmonics. They cause extra stress to railway equipment and can interfere with sensitive signalling circuits resulting in their mis-operation. Therefore, propagation of current harmonics in the catenary network and resonance phenomena have to be investigated, which requires network calculations for high frequencies. A new catenary network calculation method covers the frequency range up to 20 kHz. Experimental proof of the calculation method was done. Comparison of calculations and experimental results is given.

Keywords: network model, network calculation, catenary, track circuits, frequency converter, harmonics, electromagnetic compatibility.

1 Introduction

The development of high power semiconductors today results in the increasing use of frequency converters in railway networks. Nowadays, most rail vehicles and also some power stations are equipped with frequency converters. Such converters generate an AC-voltage by building it with pulse shaped voltages (pulse width modulation). Because of that, the resulting output voltage of the



converter consists of a main frequency voltage and harmonics (e.g. Etzeberria-Otañui *et al.* [1]).

Figure 1 shows a simple traction network. The frequency converter of a locomotive generates voltage harmonics $U_v^{(conv)}$. Due to the network impedance, voltage harmonics cause current harmonics. The v -th current harmonic $I_v^{(conv)}$ of the converter can be calculated as

$$I_v^{(conv)} = U_v^{(conv)} / Z_v, \quad (1)$$

with Z_v as input impedance of the network for the v -th harmonic seen from the converter. Current harmonics of the converter propagate into the overhead catenary. Flowing back through the rails to the converter, current harmonics can affect track circuits.

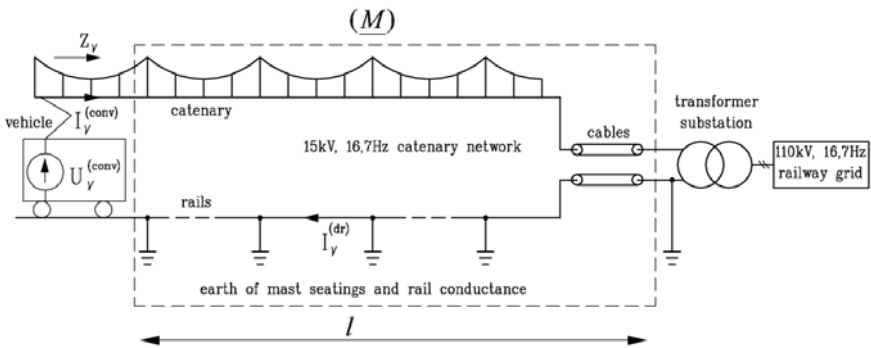


Figure 1: Traction network.

Track circuits are used for signalling purposes and are one of the most important parts of a railway system. They serve to detect the presence of trains on a particular track section. Track circuits of German railways operate at frequencies of 42 and 100 Hz as well as at audio frequencies from 4,75 kHz up to 16,8 kHz. A track circuit uses rails as part of a circuit.

Returning back to the frequency converter, current harmonics flow also through the rails. Therefore, if a harmonic current is in the same frequency range as the operating frequency of the track circuit, it can interfere the track circuit and lead to a mis-operation of the detection relay [2].

Propagating along the catenary, current harmonics of the converter can be, depending on the network conditions, either damped or amplified. Because considered track circuits are distant from the converter place, the propagation of current harmonics along the catenary network must be investigated. The current $I_v^{(dr)}$ at the location of the detection relay can be calculated using a location-dependent damping ratio D_v [3]



$$I_v^{(dr)} = I_v^{(conv)} \cdot D_v = \frac{U_v^{(conv)}}{Z_v} \cdot D_v. \quad (2)$$

Magnitudes of voltage harmonics $U_v^{(conv)}$ depend on a frequency converter type and the height of mains voltage. They can be calculated or measured for a given converter type. Thus, the influence of the converter harmonics on the track circuits is determined mainly by the damping ratio and impedance which are heavily dependent on the harmonic's frequency and network parameters such as catenary length, length of feeding cables in substations, type of catenary, location of vehicles, substations and track circuits and so on [4].

Thus, it is very difficult to make a general statement about the electromagnetic compatibility of a certain converter type basing on a few measurements. Because of that we propose to check interference limits using the model shown in figure 2. The model consists of a network model and a track circuit model. This paper is focused on the network model. Benefits and limits of known network models are discussed. Then, a new network model developed by the authors is presented. A comparative analysis of the discussed models is given. The theoretical model has been verified with the help of measurements. The results are given in the paper.

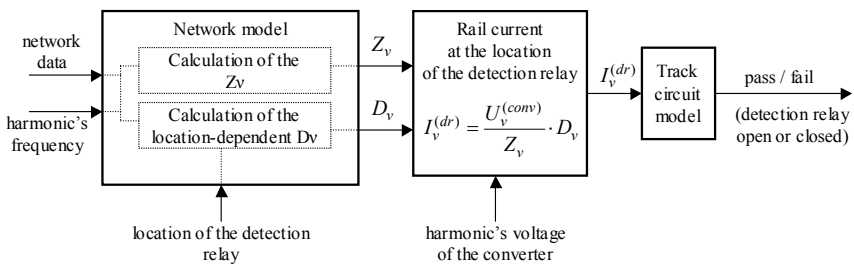


Figure 2: Model for checking interference limits.

2 Overview of known network models

2.1 General requirements to a network model

In order to check electromagnetic compatibility of frequency converters and track circuits with the help of the model figure 2, the network model must be applicable for long non-homogeneous catenary networks with railway stations, substations, tunnels and so on. A catenary can consist of many wires and be configured in any manner. The network model must be applicable for high frequencies and be able to calculate rail currents and their propagation along the track. Below, an overview of the known network models is given.



2.2 Network simulation model

This approach consists in building the model of the network circuit in a simulation program such as Simulink of the MathWorks Company or Simplorer of the Ansoft Corporation. All nodes of the original network circuit must be replicated in the simulation model. In railway catenary networks with return current conduction through the soil, the earth points of catenary masts are to be considered as nodes. Because masts are usually installed every 40-80 m along the track, the number of nodes to be replicated in the simulation model is very high. For long tracks with many nodes the simulation model becomes so complex and requires so much computation costs, that it becomes no more applicable. But a simplified replication of the network in the simulation model would result in methodical errors.

2.3 Two-wire transmission line model

This network model is used for calculation of two-wire systems. The interrelation of input and output voltages and currents U_{in} , I_{in} , U_{out} , I_{out} of a two-wire system of the length l is described with the well-known transmission line equations

$$\begin{cases} U_{in} = U_{out} \cdot ch(\gamma \cdot l) + I_{out} \cdot Z_w \cdot sh(\gamma \cdot l) \\ I_{in} = U_{out} \cdot sh(\gamma \cdot l) / Z_w + I_{out} \cdot ch(\gamma \cdot l) \end{cases} \quad (3)$$

with γ and Z_w as propagation factor and wave impedance of the two-wire system respectively.

Railway catenary networks, in which the soil is used as a return current conductor, must be considered as many-wire systems consisting of at least three conductors (contact line, rail and soil). In order to apply the line equations, the number of conductors must be reduced to two. For this purpose, several feeding current conductors (contact and carrying wires) and return current conductors (soil, rails and return current wires) must be consolidated [5]. The two-wire system, obtained in that way, can be calculated with the line equations (3).

However, a consolidation of wires without methodical errors is possible only if a current distribution in single wires is balanced and remains constant along the railway track. This requirement is fulfilled only for homogeneous tracks. In the vicinity of track boundaries and non-homogeneous transition areas (substations, tunnels, railway stations, rail vehicles and so on) a current distribution in rails and soil is not balanced [6]. In this case, the consolidation of conductors leads to methodical errors. Thus, the model does not meet the above-mentioned requirements.

2.4 Network model with multipoles

The main idea of this calculation method is to represent the track with a multipole described with its transmission matrix ($\underline{M}$) as shown in figure 1. In this case, the following matrix equation takes place



$$(\underline{UI}_{out}) = (\underline{M}) \times (\underline{UI}_{in}) . \quad (4)$$

$(\underline{UI}_{in})$ and $(\underline{UI}_{out})$ are column vectors of the input and output voltages and currents of the multipole. The elements of the matrix $(\underline{M})$ can be calculated as described in [6] and [7]. Additionally, capacitive couplings of conductors must be considered for high frequencies.

The matrix equation (4) can be solved together with boundary conditions, which are defined by node equations for voltages and currents on the ends of the track. Currents and voltages of all conductors along the track can be computed with the help of one of the classical circuit calculation methods.

The described multipolar calculation method was implemented in a MATLAB program. Our investigations have shown, that a numeric network computation with this method is possible only for low frequencies and short-length tracks. The results computed for 16,8 kHz were no more plausible for a track longer than approximately 5 km. The reason for errors is the limited accuracy of representation of numeric data in the computer.

In general, the sensitivity of the solution of a system of linear equations to errors in the data can be estimated with the help of the reciprocal condition number $RCOND$ of the matrix (M) . The reciprocal condition number can be computed for example with a *MATLAB*-Function $RCOND(M)$ and can take on values between 0 and 1. When $RCOND(M)$ is near 1, the matrix is said to be well-conditioned. For a well-conditioned matrix small changes in the data always result in small changes in the solution. When $RCOND(M)$ is near 0, the matrix is called ill-conditioned. In this case small changes in the data (for example due to the limited accuracy of their numeric representation in the computer) can lead to very large changes in the solution. Thus, the solution of the system with the ill-conditioned matrix is not accurate. Our investigations for catenary networks have shown, that the reciprocal condition number tends to zero with growing track length as well as with the frequency. Thus, the matrix (M) becomes ill-conditioned beginning from the certain track length. This results in big errors of the numerical solution of eqn (4). Therefore, this calculation method is limited to short-length tracks and low frequencies.

2.5 Summary of the overview

The overview of the known network models and their calculation methods has shown, that some models (for example, two-wire transmission line model) reproduce the real network with essential simplifications resulting in methodical errors. Other models, in which simplifications like consolidation of conductors are not mandatory, require too much computation costs (network simulations) or they cannot be calculated for long tracks and high frequencies due to numeric errors (multipolar calculation method [7]).

Thus, there is currently no model with an appropriate calculation method, which meets the requirements in full measure.



3 New multipolar calculation method

The patented method for calculation of the multipolar network model has been developed by the authors. The main idea of the new method is to reduce the numeric errors of the known multipolar calculation method by improving the reciprocal condition number of the transmission matrix of the network. The improvement of this number is achieved by increasing the matrix order. The developed multipolar calculation method enables computations for long tracks and covers the frequency range up to 20 kHz.

Consider the case when the matrix $(\underline{M})$ is ill-conditioned, i.e. $RCOND(\underline{M}) \approx 0$ so that eqn (4) cannot be solved directly due to numeric errors. Thus, the common multipolar method is not applicable.

In the proposed calculation method the whole track of the length l is subdivided into n sections of the lengths $l_1, l_2, \dots, l_n$ (figure 3). Each section can include multiple spans, masts, vehicles and so on and is represented as a multipole with its square transmission matrix $(\underline{M}_i)$, $i=1..n$. For voltages and currents at the beginning of the track are related to those at the end of the track by the equation

$$(\underline{UI}_{out}) = (\underline{M}_n) \times \dots \times (\underline{M}_1) \times (\underline{UI}_{in}) = (\underline{M}) \times (\underline{UI}_{in}). \quad (5)$$

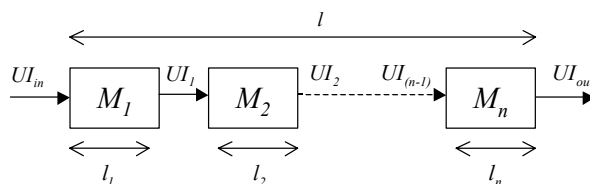


Figure 3: Subdivision of the track in sections.

The following matrix equations take place for boundaries of each section

$$\begin{cases} 0 = (\underline{M}_1) \times (\underline{UI}_{in}) - (\underline{UI}_2) \\ \vdots \\ 0 = (\underline{M}_n) \times (\underline{UI}_{n-1}) - (\underline{UI}_{out}) \end{cases}, \quad (6)$$

with $(\underline{UI}_1) \dots (\underline{UI}_{n-1})$ as matrices of input and output voltages and currents of the corresponding sections.

The subdivision in sections is to be made in such a way, that the reciprocal condition number of the matrix of each section is big enough to solve each of the equations (6) with a given accuracy [8].

Boundary conditions on the ends of the track can be expressed in a matrix form



$$(\underline{B}_{in}) = (\underline{A}_{in}) \times (\underline{UI}_{in}), \quad (\underline{B}_{out}) = (\underline{A}_{out}) \times (\underline{UI}_{out}), \quad (7)$$

with $(\underline{B}_{in})$, $(\underline{B}_{out})$ as column vectors and $(\underline{A}_{in})$, $(\underline{A}_{out})$ as matrices of the boundary conditions.

The eqns (6) and (7) can be consolidated together into a single matrix equation

$$\begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ \underline{B}_{in} \\ \underline{B}_{out} \end{pmatrix} = \begin{pmatrix} \underline{M}_1 & -I & \dots & 0 & 0 \\ 0 & \underline{M}_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \underline{M}_n & -I \\ \underline{A}_{in} & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & \underline{A}_{out} \end{pmatrix} \times \begin{pmatrix} \underline{UI}_{in} \\ \underline{UI}_1 \\ \underline{UI}_2 \\ \dots \\ \underline{UI}_{out} \end{pmatrix}, \quad (8)$$

with (I) as identity matrix of the same order as $(\underline{M}_1) \dots (\underline{M}_n)$.

With the help of matrix theory [9] it can be shown, that the matrix in eqn (8) is better conditioned than the matrix $(\underline{M})$ in eqn (4). Thus, errors of the numeric solution of eqn (8) are also smaller. They can be kept within desired tolerance limits by defining the total number of sections and the length of the each section.

The accuracy of the numeric solution increases with growing number of subdivisions and reducing lengths of the sections. But the order of the matrix in eqn (8) and therefore computation costs grow as well. Thus, the subdivision should be performed in such a way that on the one hand the number of sections is as low as possible. On the other hand, the solution of eqns (6) must be accurate enough for each section, i.e. the reciprocal condition number of the each matrix $(\underline{M}_i)$ must be big enough. In order to achieve the optimal subdivision of the track, a special iterative algorithm described in [8] should be used.

4 Comparison of calculation methods

In order to compare different network calculation methods, they have been applied to the double-track network (figure 1). A 1,3-km long four-track railway station, which is ca. 8 km distant from the frequency converter, was considered.

Figure 4 shows the rail current calculated for 16,8-kHz harmonic with three methods. The 16,8-kHz harmonic voltage of the converter was taken $U_{16,8 \text{ kHz}}^{(conv)} = 1\text{V}$. One can see, that the results from the common multipolar method become unfeasible from 5,5 km due to numeric errors. In contrast, the new multipolar calculation method provides plausible results along the whole track. The results obtained with the two-wire transmission line model diverge from those of the new multipolar method. This is to lead back to the methodical errors due to a non-balanced current distribution between rails and soil in the near of railway station and track ends.



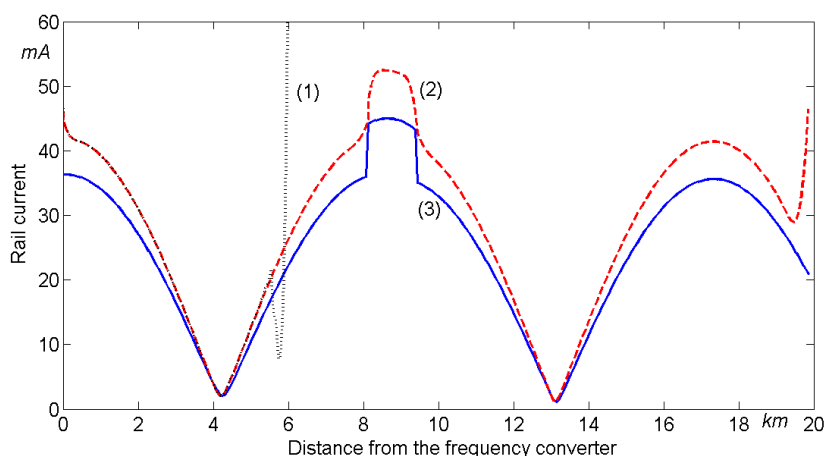


Figure 4: Rail current of the 16,8-kHz harmonic, calculated with the common multipolar method (1), new multipolar method (2), and the two-wire transmission line method (3).

5 Comparison of calculation and measurement results

In order to check the correctness of the new calculation method, measurements were done on a 25 km long single track suburban line consisting of two rails of the type S54, a carrying wire of the type Ri100 and a traction wire of the type Bz50. The supply voltage of 15 kV of the overhead catenary was powered off during the measurements. By means of a wave generator and a power amplifier the one end of the catenary was fed with a sine voltage $U_{in} = 200$ V with frequencies from 16,7 Hz up to 18 kHz. The other end of the catenary was short-circuited to the rail through the measuring resistor of 1 Ω . The feeding current I_{in} and the current I_K over the measuring resistor at the short-circuit place were measured. With the measured values the track impedance $Z = U_{in} / I_{in}$ and the damping ratio $D_K = I_K / I_{in}$ were computed.

Before doing measurements, the frequency-dependent impedance and the damping ratio have been calculated with the help of the network model. Measure and calculation results are shown in figure 5 (curves 1 and 2).

The similar character of the measured and calculated dependencies can be noted. However, the calculated curves are somewhat stretched along the frequency axis. Trying to find the reason for the discrepancy the authors have established, that mast insulators influence the results due to capacitances between their caps. According to a rough guess of the insulator capacitance (Behmann and Schäfer [10]), the value of approximately 40 pF pro capacitor could be considered. A sidearm of each mast consists of an armature and two to three insulators. Moreover, the armature parts staying under voltage build a



capacitance to the ground. Thus, the total sidearm capacitance can be guessed to be 120 pF per mast. The capacitance of the overhead catenary without taking into account sidearm capacitances was calculated as 10,4 nF/km. With 15 to 25 masts per kilometre, the additional capacitance due to the mast sidearms makes up 15 to 25% of the total capacitance of the catenary and therefore can not be neglected in the network model. The results obtained with the improved model show much better agreement with the measurements (curve 3 in figure 5).

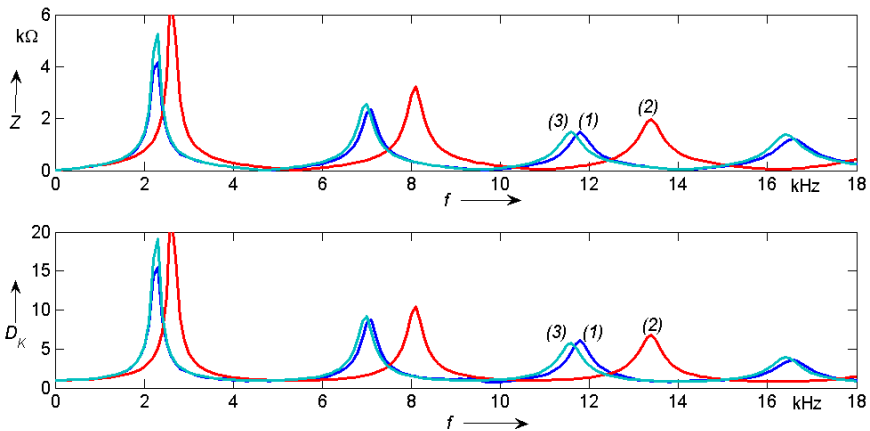


Figure 5: Calculated and measured catenary impedance and current damping ratio. (1) - measurements, (2) - calculations before measuring neglecting sidearm capacitances, (3) - calculations with the improved model taking into account sidearm capacitances.

6 Conclusions

In the presented new multipolar calculation method numeric errors of the common multipolar method have been reduced. This allows network computations for long, non-homogeneous traction networks for high frequencies. The method can be used to assess the maximal expected harmonic current emitted by a frequency converter of a vehicle or a station. Obtained results can be used for checking the electromagnetic compatibility of frequency converters and track circuits. Good agreement between calculation and measurement results testifies the correctness of the calculation method.

References

- [1] Etxeberria-Otadui, I., Bacha, S., Bachry, A., Styczynski, Z., Fracchia, M. & Stuart, M., Power electronic converters for railway infrastructure upgrading: advantages and potential risks. *Proc. of the 2nd Int. Conf. on Critical Infrastructures*. Grenoble (France), October 25-27, 2004.



- [2] ESCARV, Electrical System Compatibility for Advanced Rail Vehicles. *ESCARV Final Synthesis Report*, BRITE EURAM III - Industrial and Materials Technology, 2001.
- [3] Xie, J., Zynovchenko, A., Jank, S. & Klier, F., Calculation of high frequency harmonics in catenary networks. *eb - Elektrische Bahnen*, (103) 2005, H. 6, pp. 286-290.
- [4] Zynovchenko, A., Xie, J., Jank, S. & Klier, F., Resonance phenomena and propagation of frequency converter harmonics in the catenary of railways with single-phase A.C. *Proc. of EPE'2005 11th European Conf. on Power Electronics and Applications*. Dresden (Germany), September 2005.
- [5] Putz, R., Über Streckenwiderstände und Gleisströme bei Einphasenbahnen. *eb - Elektrische Bahnen*, 20 (1944), H., pp.74-92.
- [6] Behrends, D., Brodkorb, A. & Hofmann, G., Calculation of catenary impedance values. *eb - Elektrische Bahnen*, 92 (1994), vol.4, pp. 117-122.
- [7] Kontcha, A., Multipolar method for calculations in multi-conductor contact line systems for railways with single-phase A.C. *eb - Elektrische Bahnen*, 94 (1996), vol.4, pp. 97-102.
- [8] *Method for calculation of electric currents and voltages of high frequencies in catenary networks of electric railways*, German patent Nr. 102005006336.5, 2005.
- [9] Golub, G. H., & van Loan, C. F., *Matrix Computations*, The Johns Hopkins University Press; 3-rd edition, 1996.
- [10] Behmann, U. & Schäfer, A., Effects from catenary admittance on results of short-circuit tests. *eb - Elektrische Bahnen* (92) 1994, H. 9, S. 260-266.



A numerical algorithm for run-curve optimization of trains considering a DC feeding circuit

H. Ko & M. Miyatake

*Department of Electrical and Electronics Engineering,
Sophia University, Japan*

Abstract

An algorithm optimizing total energy consumption of multiple train operation considering a DC feeding circuit is investigated in this paper. Our mathematical formulation includes several characteristics of trains which depend on feeding voltage. It makes it possible to give detailed consideration to an energy-saving operation. It is especially important for us to be able to discuss the influence of squeezing control of regenerating current and feeding loss. We constructed the optimizing algorithm based on the gradient method applicable to large-scale problems for future works. Several numerical examples are demonstrated to verify the reliability and validity of the proposed method. Every optimisation result is obtained within a minute.

Keywords: train operation control, optimization, energy-saving, DC feeding system.

1 Introduction

In recent years, regenerative energy of trains with electrical regenerative braking system plays an important role in reducing whole energy consumption in DC railway system. However, there are several difficulties in collecting regenerative energy efficiently. One of these difficulties is that the regenerating ability and amount of regenerating energy depend closely on feeding voltage of a train. For example, the maximum amount of the electrical braking torque is limited by feeding voltage, and feeding loss also relies on feeding voltage. In addition, it is indispensable to discuss influence of squeezing control of regenerative current for preventing overvoltage of feeder. The other difficulty in collecting



regenerative energy is preparing external load for absorbing regenerating energy, *e.g.* the other accelerating trains. This consideration conducts discussion for energy-saving operation towards an idea that acceleration/deceleration timings of several trains are important to reuse regenerative energy efficiently, in other words, optimizing several trains operation is required.

We already realized to solve the kinetic energy minimization problem of a train in previous works [1–3]. However, we did not consider the effect of energy exchange among multiple trains, because of difficulties in mathematical formulation and developing numerical algorithm.

In this paper, we formulate the energy-saving operating problem of several trains considering DC feeding system as an optimal control problem. The proposed formulation includes equality constraints of circuit equations of a feeding system. We also propose a simplified parallel numerical optimizing algorithm based on the traditional gradient method step by step. We verify the reliability and validity of the proposed method by some numerical examples about two trains operation.

2 Mathematical formulation

In this section, the energy-saving operation problem of trains is described as an optimal control problem. The variable vector $\zeta(t)$ is used for whole state vectors of system (position and speed of trains, node voltages of a feeding circuit, control inputs, and so on) in this paper.

2.1 Objective functional

The objective functional J is given as the total electrical energy consumption at substations in eqn (1).

$$J = \int_0^T \sum_{m=1}^M E_m(t) I_m(\zeta(t), t) dt \quad (1)$$

Here, T is total time, M is the number of substations. Substation source voltage E_m and substation load current I_m are defined as time-dependent function, load current I_m relies on system variables ζ .

In previous works, objective functional was chosen as total kinetic energy consumption, under a given feeding voltage variation [2, 3]. On the contrary, the objective functional J is preferable to be formulated as eqn (1) in our case, because we can consider feeding loss and effect of squeezing control with given functional eqn (1).

2.2 Trains

We formulate states and restrictions on train operation as state equations and inequality constraints in an optimal control problem, respectively. State



equations and constraints of j -th train are given as eqn (2) using position x_j , speed v_j , feeding voltage e_j , departure time t_{j0} , arrival time t_{jf} , departure position x_{j0} , arrival position x_{jf} , speed limit v_{jmax} , maximum acceleration with electrical torque f_{jmax} , maximum deceleration with electrical and mechanical blending torque f_{jmin} , and load current i_j of a train.

$$\begin{aligned} \dot{x}_j &= \begin{cases} v_j & t_{j0} \leq t \leq t_{jf} \\ 0 & \text{others} \end{cases} \tag{2} \\ \dot{v}_j &= \begin{cases} f_j(u_j, e_j, v_j) - r_j(x_j, v_j) & t_{j0} \leq t \leq t_{jf} \\ 0 & \text{others} \end{cases} \tag{3} \\ x(t_{j0}) &= x_{j0}, \quad v(t_{j0}) = 0 \tag{4} \\ x(t_{jf}) &= x_{jf}, \quad v(t_{jf}) = 0 \tag{5} \\ f_{jmin} &\leq f_j(u_j, e_j, v_j) \leq f_{jmax}(e_j, v_j) \tag{6} \\ v_j &\leq v_{jmax}(x_j) \tag{7} \\ i_j &= i_j(u_j, e_j, v_j) \tag{8} \end{aligned}$$

Eqn (2) and (3) are kinetic equations, eqn (4) and (5) are boundary conditions at departure/arrival time. Here, function f_j describes acceleration/deceleration and u_j is control input defined as table 1. Torque and speed limit of a train is described as eqn (6) and (7). Maximum acceleration f_{jmax} varies according to train speed and catenary voltage (please see figure 2(a)). Function r_j consists of running and grade resistance. Eqn (8) means load current of a train depends on kinetic power and feeding voltage. It implies that we can consider influence of squeezing control and limitation of regenerative electrical torque, too. We also assume that the mechanical braking torque is only used when the required braking torque exceeds the maximum electrical braking torque for energy-saving operation.

Table 1: Definition of control input u .

$u=-1$	maximum deceleration
$u=-1$ to 0	deceleration, the amount of brake torque is proportional to $-u$
$u=0$	coasting
$u=0$ to 1	acceleration, the amount of motive torque is proportional to u
$u=1$	maximum acceleration

2.3 Feeding circuit

We regard circuit equations of a feeding system as equality constraints in the optimal control problem shown as eqn (9) except for transient responses in and under millisecond region.

$$C_l(\zeta(t), t) = 0 \quad (l = 1, \dots, P) \quad (9)$$

Here, P is the number of circuit equations. In general, circuit equations of a feeding system are too complicated to solve analytically. Thus, a numerical optimization algorithm which does not require analytically solvable equality constraints is suitable for the optimal control problem of trains considering feeding system.

2.4 Representation as optimal control problem

Finally, we formulate the energy-saving operating problem of trains as an optimal control problem (10) with equality and inequality constraints considering previous sections 2.1~2.3. The variable N indicates the number of trains.

$$\min J, \quad \text{subj. to } (2), \dots, (9) \quad (10)$$

$$\{u_j\}_{j=1, \dots, N}$$

We assume every function in eqn (10) is differentiable using the mollifier if needed. This optimal control problem (10) could be a large-scale problem.

3 Numerical algorithm

In this section, a numerical algorithm for the formulated problem (10) is derived using some transformation techniques and the optimality condition. The variables without lower position indexes mean sets of the all elements corresponding variables (for example, $u = \{u_1, u_2, \dots\}$).

3.1 Optimality and fundamental algorithm

The formulated optimal control problem (10) is so called a two-point boundary value problem which has constraints on initial state variables (4) and terminal states (5). It is generally considered that an optimal control problem of this class meets any difficulties to solve numerically. A transformation technique to initial-value problem is employed to avoid these difficulties. The modified objective function is derived as eqn (11).

$$J' = J + \sum_{j=1}^N \left\{ p_{j1} (x(t_{jf}) - x_{jf})^2 + p_{j2} v(t_{jf})^2 \right\} \quad (11)$$

Here, p_{j1} and p_{j2} are designed penalty parameters adjusted by trial and error task, but it is easy to perform an automated adjustment algorithm [6].

The other transformation from constrained input problem to unconstrained one employing slack control input u' (such as $u = \sin(u')$) is also used to simplify the formulated optimal control problem (10). Finally, the simplified optimal



operating problem (12) is derived from these transformation techniques and assumption that the solution of circuit equation (9) is unique.

$$\begin{aligned} & \min_{\{u'_j\}_j, \{e_l\}_l} J' \\ \text{subj. to} \quad & \dot{\xi}_j = g_j(u'_j, \xi_j, t), \quad \xi_j(t_{j0}) = \xi_{j0} \quad (j=1, \dots, N) \quad (12) \\ & C_l(\zeta, t) = 0 \quad (l=1, \dots, P) \end{aligned}$$

Here, the state vector $\xi_j = (x_j, v_j)^T$ and the vector function g_j are introduced to simplify the expression. The vector function g_j consists of the right hand side of eqn (2) and (3). The variable e_l is assumed as quasi-control-input to solve the circuit equations (9) numerically.

The optimality condition of problem (12) is given as follows [4].

$$H_{u'_j} \equiv \partial H / \partial u'_j = 0 \quad (13)$$

$$H_{e_l} \equiv \partial H / \partial e_l = 0 \quad (14)$$

The Hamiltonian H is defined as eqn (15) and (16) introducing variable vector λ_j, π_l .

$$H = \sum_{m=1}^M E_m I_m + \sum_{j=1}^N \lambda_j^T g_j + \sum_{l=1}^P \pi_l C_l \quad (15)$$

$$\dot{\lambda}_j = -\partial H / \partial \xi_j \quad (16)$$

It is a well-known result as one of the optimal control theories that the function $H_{u'}$ presents gradient function which gives the steepest ascending direction. Thus, the numerical optimization algorithm for problem (12) is given as follows considering that the vector π_l is found as solution of equality (14).

<Algorithm I>

- give the initial estimation u^0, e^0 , and set $k=1$.
- solve ODE(Ordinal Differential Equation)s (2),(3) and (4) using u^k, e^k , and obtain ζ^k .
- solve ODE (16) and equalities (14), and obtain λ^k, π^k .
- calculate gradient function by (13), using $\zeta^k, \lambda^k, \pi^k$.
- execute line search toward the search direction (steepest gradient direction, conjugate gradient direction, and so on), and set the solution u^{k+1} .
- solve equality (9), and set the solution as e^{k+1} .
- if the difference between e^k and e^{k+1} is adequately small, the procedure terminates, or else set the $k=k+1$ and return to step (b).



This algorithm is based on the traditional gradient method for unconstrained optimal control problem.

3.2 Parallel algorithm

As mentioned in section 2.4, the formulated optimal train operating problem (10) (or eqn (12)) could be a large-scale problem. Parallel algorithms are valuable because these are faster to perform large computing task. A parallel algorithm for the formulated problem is prepared employing Kirchhoff's Current Law (KCL) as follows. The objective functional J is expanded with KCL as eqn (17).

$$J = \int_0^T \left(\sum_{j=1}^N \alpha_j i_j + \sum_s \alpha_s i_s \right) dt \quad (17)$$

Here, α_j and α_s are the coefficients derived by expansion, and the variable i_s is a current without load currents of trains. Thus, the objective functional J is divided for every train shown as eqn (18) distributing the terms which do not involve load current properly.

$$\begin{aligned} J &= \sum_{j=1}^N \int_0^T \left(\alpha_j i_j + \sum_q \sigma_q \right) dt \\ &\equiv \sum_{j=1}^N U_j \end{aligned} \quad (18)$$

Here, variable σ_q is the term belonging to U_j .

Then, a parallel algorithm for the formulated problem (12) is performed as the following sequence.

<Algorithm II>

- (a) give the initial estimation u^0, e^0 , and set $k=1$.
- (b) solve ODEs (2),(3) and (4) using u^k, e^k , and obtain ζ^k .
- (c) solve ODE (16) and equalities (14), and obtain λ^k, π^k .
- (d) calculate gradient function for every train, using $\zeta^k, \lambda^k, \pi^k$.
- (e) execute line search in the search direction for every train, and set the solution u^{k+1} .
- (f) solve equality (9), and set the solution as e^{k+1} .
- (g) if the difference between e^k and e^{k+1} is adequately small, the procedure terminates, or else set the $k=k+1$ and return to step (b).

In this algorithm, step (d) and (e) could be proceeded simultaneously for every train. Namely, the original optimal control problem is distributed for every train and interactions between trains are mainly informed by e, π .



3.3 Simplified algorithm

The parallel algorithm shown in subsection 3.2 could be valuable for a large-scale problem formulated as eqn (10). However, this algorithm would still have some difficulties. The variable vector π is found as solution of eqn (14) mentioned above. But, the equality (14) can become numerically unstable or its solutions are not unique because of the complexity of the feeding circuit such as characteristics of diode rectifier. For this reason, we propose the following intuitive approximated method.

<Algorithm III>

- (a) give the initial estimation u^0, e^0 , and set $k=1$.
- (b) solve ODEs (2),(3) and (4) using u^k, e^k , and obtain ζ^k .
- (c) solve ODE (16) neglecting the terms including π^k , and obtain λ^k .
- (d) calculate gradient function for every train neglecting the terms including π^k , using ζ^k, λ^k .
- (e) execute line search in the search direction for every train, and set the solution u^{k+1} .
- (f) solve equality (9), and set the solution as e^{k+1} .
- (g) if the difference between e^k and e^{k+1} is adequately small, the procedure terminates, or else set the $k=k+1$ and return to step (b).

This simplified algorithm does not satisfy the optimality condition of the original problem (10), but step (b), (c), (d) and (e) could be proceeded simultaneously for every train, and the solution of equality (14) is not required. Namely, this algorithm is more suitable for large-scale problems and numerically stable than Algorithm II. Algorithm III is performed based on the assumption that interactions between trains are mainly informed with feeding voltage e .

4 Numerical examples

In this section, three numerical examples of two-train-operation in DC feeding system shown as figure 1 are demonstrated using Algorithm III. Please refer [6] for comparison between Algorithm II and III. In the following examples, conjugate gradient direction is employed as a search direction in line search, and Clipping-Off algorithm [5] is used to deal with inequality constraints on control inputs instead of slack control inputs. It does not cause any essential difference from our proposed method. The C++ language on an Intel Celeron 1.4GHz machine is employed for demonstrations.

The characteristics of motive and brake torque, and squeezing control are presented in figure 2. The supply voltage of substations E_s , internal resistance of substations R_s and line resistance R_l is given as 1500[V], 0.05[Ω] and 0.04[mΩ/m], respectively. All trains transfer 2180[m] in 130[sec] between stations. The first train departs from the position $x=0$ [m] at time $t=0$ [sec], and the second train departs from the position $x=2180$ [m] at time $t=t_s=90, 65$ or 40[sec]. The departing time of the second train is changed in order to change the phase of two trains.



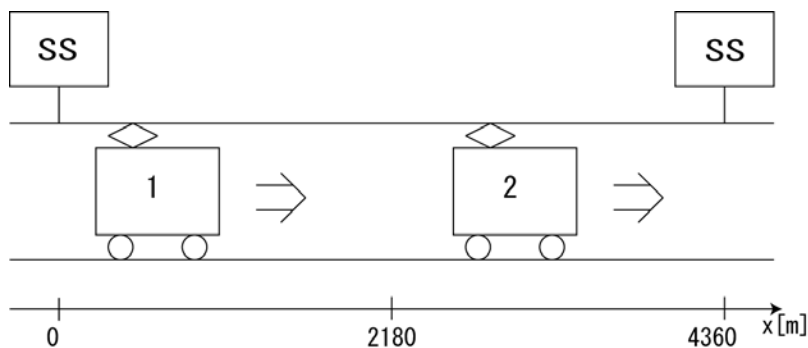


Figure 1: Schematic model of numerical examples.

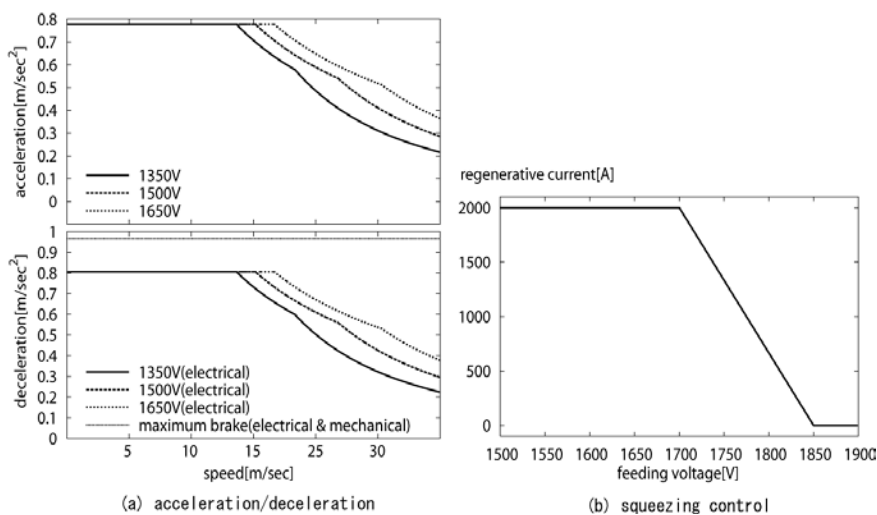


Figure 2: Characteristics of a train.

The optimization results of these three cases are depicted in figure 3, in which horizontal axis is chosen as the normalized time $\tau=t/T$. The optimal energy consumptions are found as (a) 244.1[MJ]($t_s=90$), (b) 280.4[MJ]($t_s=65$), (c) 329.1[MJ]($t_s=40$). Each optimizing procedure is converged in less than 1 minute.

All the results indicate that the first train acts like an energy storage device. The first train can collect electrical energy with comparatively high efficient near the departure position, on the contrary, greater feeding loss occurs when the second train starts. Thus, the first train accelerates up to higher speed to store electrical energy as kinetic energy in advance, and supplies the stored energy when the second train accelerates in order to reduce feeding loss. The amount of braking torque of the first train is adjusted to minimize the effect of squeezing control, if the maximum braking torque were employed in higher speed area, the amount of regenerative energy is cut down as the following squeezing

characteristics shown in figure 2 (b). Comparison task among the optimization results with three different departure time of the second train shows when the proper time is to use regenerative energy most efficiently for energy-saving operation. If the departure time of the second train is preferable ($t_s=90$), braking interval of the first train and accelerating interval of the second train is the same. In other cases ($t_s=65$ or 40), regenerative energy of the first train is used for accelerating stage of the second train in comparatively high-speed area because electrical power demand and feeding loss become greater at this point.

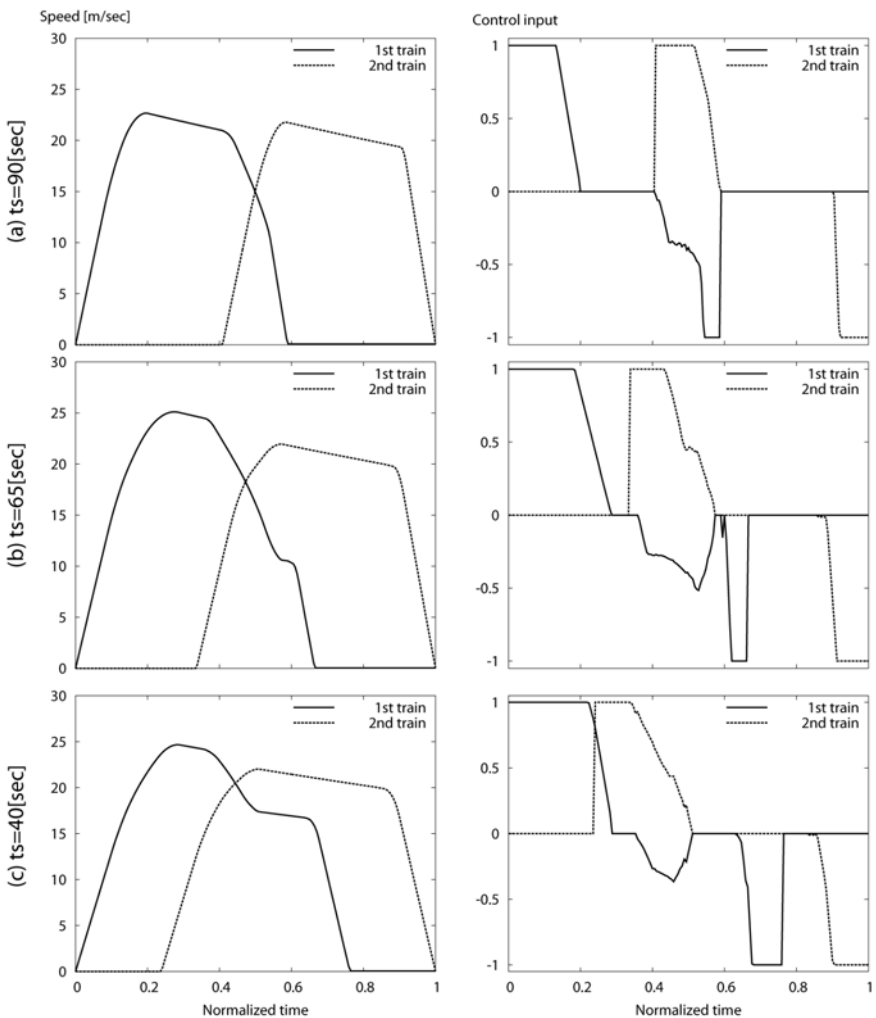


Figure 3: Optimization results with different departing phase.

These considerations for energy-saving operation were already discussed in our previous works. However, we only mentioned qualitative analysis, not based



on the quantitative optimal solutions. Our newly proposed formulation and numerical method could overcome these insufficient discussions and make important role at this point.

5 Conclusion

The optimal operating problem of multiple trains considering DC feeding system is formulated and the simplified approximated numerical algorithm is performed. We can find that the proposed method has enough reliability and performance from numerical examples of two trains. We also show that the influence of operating phase between two trains can be evaluated quantitatively. Further validation using models with more than two trains and energy storage devices would be our future works.

References

- [1] Khmelnitsky, E., An Optimal Control Problem of Train Operation. *IEEE Trans. on Automatic Control*, **vol.45, No.7**, pp.1257-1266, 2000.
- [2] Ko, H., Miyatake, M. & Koseki, T., Application of Dynamic Programming to Optimization of Running Profile of A Train, *Comprail2004*, pp.103-112.
- [3] Ko, H. & Miyatake, M., Fundamental Study on Train Run-curve Optimization Problem Employing Gradient Method, *National Convention of IEEJ*, vol.4, pp378-379, 2005 (in Japanese).
- [4] Kanou, H., *Theory and Computational Methods in Optimization*, 1987 (in Japanese).
- [5] Quintana, V.H. & Davison, E.J., Clipping-off Gradient Algorithm to Compute Optimal Controls with Constrained Magnitude, *Int. J. Control*, **vol.20, No.2**, pp243-255.
- [6] Ko, H. & Miyatake, M., A Numerical Method for Run-curve Optimization of Trains Considering Feeding Circuit, *TER05-35*, pp31-36, 2005 (in Japanese).



Railway modelling for power quality analysis

M. Chymera, A. C. Renfrew & M. Barnes

University of Manchester, UK

Abstract

Power quality is increasingly important in the consideration of railway systems. Poor power quality affects the performance of trains and introduces problems for local distribution networks.

This paper looks at improving modelling techniques to assess the power quality of a railway system. Modelling techniques used to simulate electrified railways are adapted to simulate power quality aspects with improved computational efficiency. The model developed is used to assess the voltage regulation and harmonic distortion of a rail system, and harmonic distortion on the local distribution network.

Keywords: power quality, harmonics, electric traction, modelling.

1 Introduction

Power quality has become an increasing concern in railway systems. Poor power quality affects the performance, reliability of the railway system as well as having an effect on equipment attached to the local distribution network. The analysis of power quality on a rail system is essential, to enable the analysis of train performance and to assess the effects of a rail system on the adjacent distribution network.

Railway systems are electrically complex. The loads, trains, are constantly moving and their electrical behaviour is constantly changing. Modelling is an ideal tool to analyse the power quality of such a complex system. An ideal model would be accurate yet computationally efficient.

2 Power quality

Power qQuality has to be considered from two perspectives, the railway system and the local distribution network to which it is attached. Poor power quality on



the railway system affects performance, limits capacity and may introduce a need for additional maintenance. Distribution network operators stipulate power quality standards. Railway systems are required to meet these standards.

Power quality is divided into categories [1]. Railway systems typically suffer from poor voltage regulation and harmonic distortion. Voltage regulation affects the performance of traction units and imposes a capacity limit on the system. Harmonic distortion can cause damage to traction units, particularly motors [2].

Rail systems can cause harmonic distortion and voltage unbalance on the local distribution network. The harmonic distortion is caused by the non-linear currents drawn by the railway system and in DC systems by rectifying substations. Voltage unbalance is present in AC systems, where the rail system draws large currents from a single phase of the supply.

3 Modelling railway power quality

Modelling is required to assess the power quality of a rail system and the effect of the rail system on the local distribution network. Analysis is essential for assessing existing systems to establish the need for power quality compensation, or improvements to systems, and for the analysis of new systems or the effect of major improvements to an existing system.

Various modelling techniques exist for the analysis of railway electrical behaviour. Many models are developed to simulate train movement. Some models have been produced to analyse power quality, however these models tend to focus on single aspects of power quality, such as harmonic analysis [3] and voltage regulation [4]. It would be desirable and more effective to simulate these aspects of power quality together.

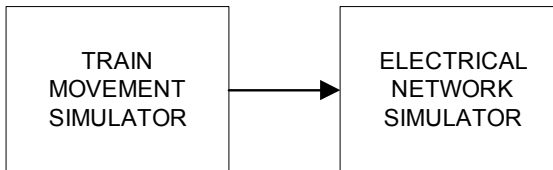


Figure 1: Rail system model.

Figure 1 shows the model developed to simulate power quality in a rail system. The train movement simulator uses input data to calculate train positions and electrical behaviour. The electrical network simulator uses the positions to calculate the line impedances, and load currents.

4 Power quality analysis

Power quality can be analysed from voltage and current waveforms. To produce voltage and current waveforms the rail system has to be analysed using differential equations. This requires short time steps and results in long computation times when performed as a time stepping transient analysis.



Assessing some of the relevant power quality aspects individually in the frequency domain may result in more efficient simulation.

4.1 Voltage regulation

Voltage regulation can be analysed by using steady-state analysis. The rail system is modelled as an electrical network and the network is analysed at each time step. Nodal voltage analysis provides an ideal tool for modelling voltage regulation.

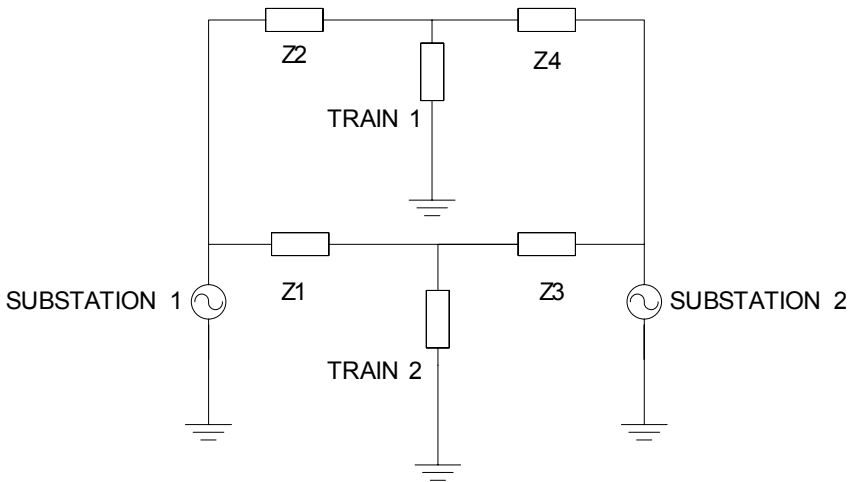


Figure 2: Electrical representation of a simple railway system.

Nodal voltage analysis uses matrices to represent the rail system. A simple rail system, Figure 2 can be represented using eqn (1). The line impedances are calculated from the train positions.

$$\begin{pmatrix} \frac{1}{Z_1} & \frac{1}{Z_3} \\ \frac{1}{Z_2} & \frac{1}{Z_4} \end{pmatrix} \begin{pmatrix} V_{\text{substation1}} \\ V_{\text{substation2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{Z_1} + \frac{1}{Z_3} & 0 \\ 0 & \frac{1}{Z_2} + \frac{1}{Z_4} \end{pmatrix} \begin{pmatrix} V_2 \\ V_3 \end{pmatrix} + \begin{pmatrix} I_{T1} \\ I_{T2} \end{pmatrix} \quad (1)$$

4.2 Harmonic distortion

Conducting frequency domain analysis and performing network analysis for each harmonic frequency offers a more efficient method of analysing waveform distortion, if superposition can be assumed. The model of each train must consider the harmonics currents drawn. This means each load produces a table of harmonics including a frequency and a magnitude.



Harmonic distortion is considered within the railway system and on the local distribution network. The harmonic distribution is simulated in 5 stages:

1. Model of the substation is configured for the railway side.
2. The magnitudes of harmonic currents drawn by the train are calculated
3. Network analysis is performed on all the frequencies present.
4. The external models of the substations are constructed.
5. Network analysis of all frequencies is performed for the distribution network.

4.2.1 Substation model – railway side

For an AC system the substation is modelled as an AC source, to represent the fundamental. If there is significant harmonic distortion on the distribution network, additional voltage sources are included in this model. The magnitude of the sources depends on the behaviour of the distribution network. The distribution network model is simplified to a static model and hence the magnitudes of the voltage sources are predetermined and stored in a table.

For a DC system, where the substation contains a transformer and a rectifier, the substation is modelled as a set of voltage sources at various frequencies. The magnitude of each voltage source is dependent on the total current drawn from the substation. The current drawn is determined when the analysis of the DC voltage is considered, after the initial network analysis is complete, the substation ripple frequency magnitudes are determined.

4.2.2 Train harmonic model

Modern trains employ power electronic drives which draw harmonic currents. The harmonics drawn depend on the type of drive system and the state of the train (acceleration and speed). Analysing the train, the harmonics drawn can be calculated. Equations can be established to determine harmonic currents drawn by the train. The train model must also consider the attenuation of harmonic currents by filters.

4.2.3 Railway side harmonic analysis

Network analysis, employing nodal voltage analysis is used to simulate the harmonics in the railway system. The equation used for the voltage regulation analysis, eqn (1) is used to perform the analysis of each harmonic. The impedances per unit length are calculated for each frequency using eqn (2). The line impedances are calculated from the train positions and the impedance per unit length.

$$Z = \sqrt{R^2 + (2\pi fL)^2} \quad (2)$$

4.2.4 Substation model – grid side

The currents drawn by the rail system from each substation at various frequencies have been determined; the substation model is used to determine the currents drawn from the grid. For an AC system the harmonics drawn from the substation are the same as those drawn by the railway system. For DC systems



the harmonics drawn are the harmonics of the system plus the harmonics drawn by the rectifier.

4.2.5 Distribution network analysis

The distribution network is modelled, with the rail system as a load, Figure 3. A matrix is constructed to represent the distribution network impedances. Using harmonic data from the sources, and transmission line impedances the network is analysed. The analysis can be performed for test systems [5] to perform general harmonic analysis of a rail system.

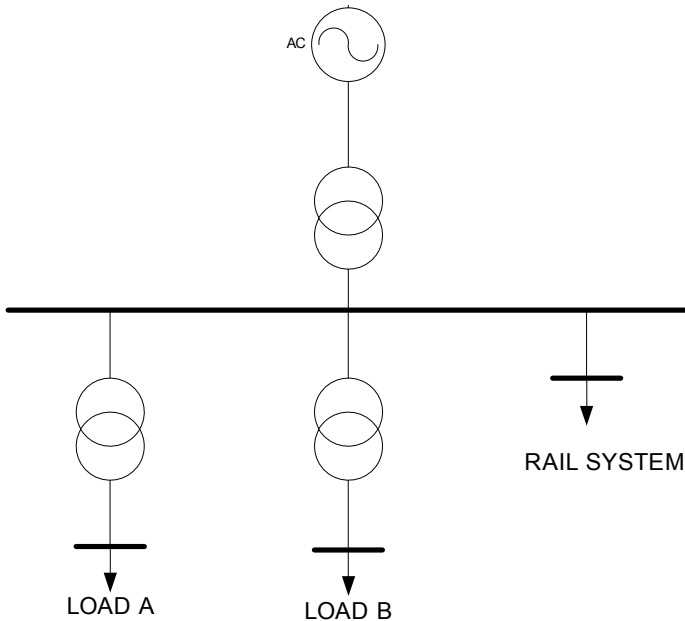


Figure 3: Simple distribution network.

5 Modelling the Blackpool Tramway

The modelling techniques were used to simulate the Blackpool Tramway, a 550V DC tram system.

5.1 Tram model

Blackpool tram system employs thyristor control trams. The chopper operates between 40Hz and 800Hz. The Blackpool Tram's use a filter to attenuate ripple, the filter characteristics are shown in Figure 4. The filter significantly attenuates ripple, and hence the ripple currents drawn by the tram are considered negligible.

The substation on the Blackpool use 12-pulse rectifiers; a 12 pulse rectifier frequency domain model is represented in Table 1.



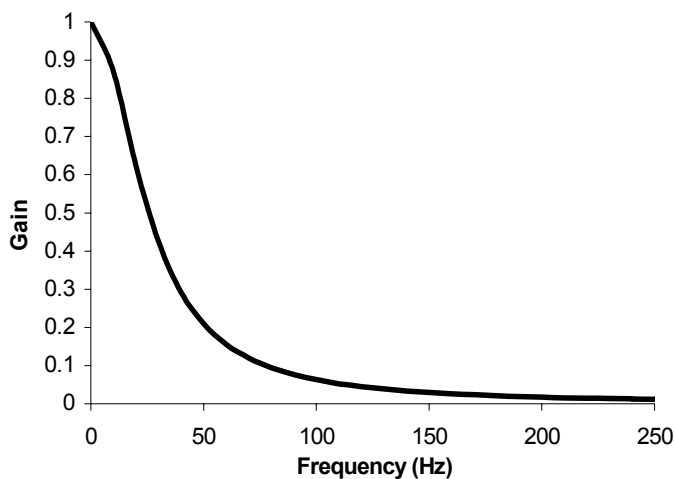


Figure 4: Blackpool Tram filter characteristic.

Table 1: Frequency domain model for substation (railway side).

Frequency/Hz	600	1200
Magnitude/A	0.014I	0.010I

Table 2: Frequency domain model of substation (distribution network side).

Frequency/Hz	550	650	1150	1250	1750	1850
Magnitude/A	0.0644I	0.0329I	0.0110I	0.0108I	0.0053I	0.0042I

The external model of the railway hence considers a linear load on a 12-pulse rectifier. The harmonic currents drawn by the rail system are dependent on the current. Table 2 represents the harmonic currents as a function of current.

6 Results

6.1 Voltage regulation

Figure 5 shows the voltage regulation curve for a tram on the Blackpool Tramway as the tram starts from rest and accelerates to 30mph, the tram then drives for 5 minutes before a short period of coasting and braking. The curve shows a voltage minimum of 537V.

6.2 Harmonic analysis

The simulation results demonstrate a constant DC voltage ripple of 1.7% within the tram system. The total harmonic current distortion at the substation is 7.6%. The Simulation results showed a negligible effect of the tram system on the Total Harmonic Voltage distortion of the local distribution network.

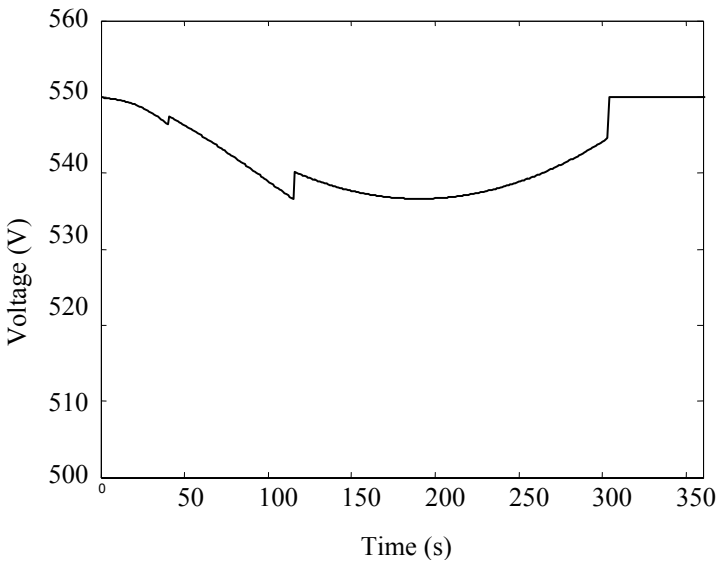


Figure 5: Voltage regulation on the Blackpool Tramway.

7 Discussion

The simulation results of the Blackpool system showed that the Blackpool tramway has reasonable voltage regulation at its current usage levels. The harmonic analysis showed a small ripple within the system of 1.7%, this figure is small enough not to cause any interference concerns. The harmonic analysis on the external systems showed that there is significant harmonic current distortion; however this does not propagate onto the distribution network due to the small relative size of the system.

8 Conclusion

The modelling techniques described provide an effective way of simulating rail system power quality. Results show the voltage regulation and harmonic distortion on the railway network and on the local distribution network. The tools can be used in the design or modification of railway schemes. The voltage regulation output is an indicator of the railways potential performance. The harmonic analysis allows the identification of any potential interference problems, particularly electromagnetic interference with signalling circuits. The harmonic analysis on the distribution network can be used to check if the grid code standards are met, and to assess the requirements for compensation.

The next stages of the work require verification of the model. Experimental results are required to assess the accuracy of the model.



Acknowledgements

The Authors would like to thank the Engineering Physical Science Research Council and HILTech Developments Ltd, who are funding the research. Thanks also to Blackpool Transportation and its Chief Electrical Engineer Peter Brown, who have provided data on the Blackpool Tramway, used to construct the models described.

References

- [1] Schlabbach, J., Blume, D. & Stephanblome, T., *Voltage Quality in Electrical Power Systems*. London: IEE, 2001.
- [2] Lee, C.Y., & Lee, W.J., Effects of nonsinusoidal voltage on the operation performance of a three-phase induction motor. *IEEE Transactions on Energy Conversion*, **14**, pp. 193-201, 1999.
- [3] Chang, G.W., Lin, H.W., & Chen, S.K., Modeling characteristics of harmonic currents generated by high-speed railway traction drive converters. *Power Delivery, IEEE Transactions on*, **19**, pp. 766-773, 2004.
- [4] Hill, R.J., & Cevik, I.H., On-line simulation of voltage regulation in autotransformer-fed AC electric railroad traction networks. *Vehicular Technology, IEEE Transactions on*, **42**, pp. 365-372, 1993.
- [5] Abu-Hashim, R., Burch, R., Chang, G., Grady, M., Gunther, E., Halpin, M., Harziadonin, C., Liu, Y., Marz, M., Ortmeier, T., Rajagopalan, V., Ranade, S., Ribeiro, P., Sim, T., & Xu, W., Test systems for harmonics modeling and simulation *IEEE Transactions on Power Delivery*, **14**, pp. 579-87, 1999.



A mixed AC/DC model for railway power systems

J. Muñoz-Riesco, E. Pilo, A. Fernandez & P. Cucala
*Instituto de Investigación Tecnológica,
E.T.S. de Ingeniería Univ. Pont. Comillas de Madrid, Spain*

Abstract

A method to study railway power supply systems with AC power distribution networks and DC power feeding is presented in this paper. The aim of this work is to obtain a DC model of the whole system in which AC networks are converted into equivalent DC ones.

Relationships between AC and DC quantities on both sides are set using the equations of multiple bridge rectifiers as described in the literature. Particular DC models for transformers and 3-phase lines are used to represent the corresponding voltage drops. Specific base quantities are used for AC and DC parts of the circuit in order to obtain a simplified circuit.

This model is used to solve a DC load flow problem. Real values of voltages, intensities and power flows can be obtained by using the proper base quantities. Additionally a correction has to be made to improve the power flow results calculated previously, due to pure resistive representation of transformers and AC lines, so a fine approximation is achieved.

Keywords: railways, AC/DC power supply system, equivalent DC model.

1 Introduction

Railways' power systems usually have an AC division (zones close to distribution networks) and a DC one (zones close to trains). This feature makes those structures difficult to deal with when an electrical study is required to be practised. Homogenization of the electrical circuit to AC or DC helps it to be solved by means of a load flow problem.

Converting the whole circuit to AC provides exact results, but it involves a great amount of variables and calculations.



Otherwise, transforming the configuration to DC does not offer so good conclusions (because of the AC part's reactance) but is easier and the calculus load is lower.

The proposal of this paper is to establish a method to solve railways' power systems with DC magnitudes minimizing the error.

2 Railway's power system description

Railway's power system is usually fed in DC power, generally in 3000V or 1500V. Figure 1 shows the typical structure in this type of arrangements. Electricity supply is done from different points of the distribution network, which will be designated *supply points*. Connected to those points through several 3-phase lines are located the *traction transformers*, which are commonly made up of a transformation stage and a rectification stage, generally consisting of multiple 12-pulse bridge rectifiers, Kaller [2]. Those *traction transformers* cause a voltage drop between the positive and negative overhead contact lines.

Electricity can be delivered from several *traction transformers* simultaneously. Sometimes, depending on the separation between these elements, a parallel feeding is done.

Railways' electrical systems are topologically similar to railways' networks; thus it involves marked radial features, Profillidis [3]; except from urban centres.

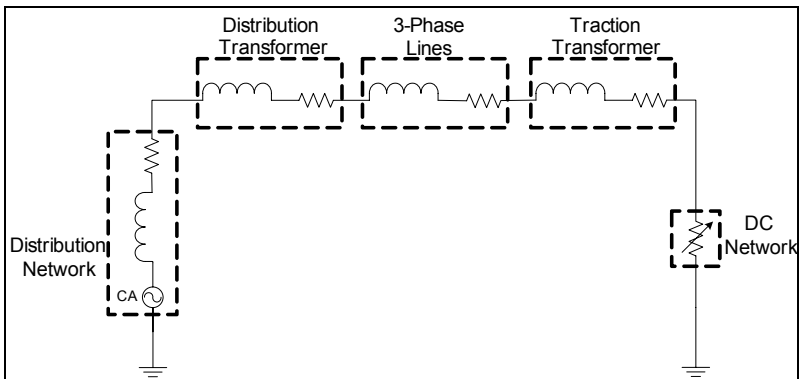


Figure 1: Railway's power system scheme.

3 Electrical model

As it has been observed previously, the AC part of the circuit will be adapted appropriately to come into being a single DC circuit. Figure 1 illustrates the group of elements, which compose the railway's power system.

The AC side can include several voltage levels: very high voltage (VHV) from the supply point to the *distribution transformer* (if it exists), high voltage (HV) in the 3-phase lines and medium voltage (MV) after the traction transformer.



The DC part has only one voltage level, which depends directly on the medium voltage rectification given by the *traction transformer*. Every train will consume energy at this voltage.

The first step is to convert every quantity in the AC part to a unitary one that must be coherent with the values of power flow and voltage in the DC fraction of the circuit.

The second step is to convert the DC part to unitary quantities consistent with those used in the AC elements.

The third step is to adjust an equivalent resistance representative of resistance and reactance values in the AC section.

Power flow base DC quantities (P_{BASE_DC}) and AC quantities ($P_{BASE\_AC\_3-PHASE}$) are identical in the whole circuit, eqn. (1), for the reason that the transmitted power does not depend on the voltage or if it is AC or DC.

$$P_{BASE\_DC} = P_{BASE\_AC\_3-PHASE} \quad (1)$$

The relationship between *traction transformer's* low voltage side base AC quantities ($U_{BASE\_L-N\_AC\_MV}$) and DC base quantities (U_{BASE_DC}), eqn. (3), is set by the rectification constant (K_{RECT}), eqn. (2), where B is the number of rectifier's bridges in series (2 in this case), Kundur [1].

$$K_{RECT} = \frac{3 \cdot \sqrt{6}}{\pi} \cdot B \quad (2)$$

$$U_{BASE\_DC} = 3 \cdot K_{RECT} \cdot U_{BASE\_L-N\_AC\_MV} \quad (3)$$

Links between high base voltage ($U_{BASE\_L-N\_AC\_HV}$) and AC base's medium voltage ($U_{BASE\_L-N\_AC\_MV}$) are fixed by transformation relationships of *traction transformers* ($K_{TRAC-TRANSF}$), eqn. (5). Relationship between very high base voltage ($U_{BASE\_L-N\_AC\_VHV}$) and high AC voltage is set by *distribution transformers* ($K_{DIST-TRANSF}$), eqn. (7).

$$\frac{U_{BASE\_L-N\_AC\_MV}}{U_{BASE\_L-N\_AC\_HV}} = K_{TRAC-TRANSF} \quad (4)$$

$$U_{BASE\_DC} = 3 \cdot K_{RECT} \cdot K_{TRAC-TRANSF} \cdot U_{BASE\_L-N\_AC\_HV} \quad (5)$$

$$\frac{U_{BASE\_L-N\_AC\_HV}}{U_{BASE\_L-N\_AC\_VHV}} = K_{DIST-TRANSF} \quad (6)$$

$$U_{BASE\_DC} = 3 \cdot K_{RECT} \cdot K_{TRAC-TRANSF} \cdot K_{DIST-TRANSF} \cdot U_{BASE\_L-N\_AC\_VHV} \quad (7)$$

The specific treatment, description and modelling each device will receive is detailed following (Kaller [2], Hill [4, 5]).

3.1 Distribution network

It is modelled by its equivalent Thévenin dipole and it is connected to the *supply point*. The equivalent resistance ($R_{DIST-NET}$) and reactance ($X_{DIST-NET}$) models the voltage drop produced in the network because of trains' energy consumption.

The value of the voltage, reactance and resistance will be usually specified in total values so they must be converted to unitary quantities ($r_{DIST-NET}$ and $x_{DIST-NET}$) of the AC zone, making use of base impedance ($Z_{BASE\_AC\_VHV}$), eqns. (9) and (10).



$$Z_{BASE\_AC\_VHV} = 3 \cdot \frac{U_{BASE\_L-N\_AC\_VHV}^2}{P_{BASE\_AC\_3-PHASE}} \quad (8)$$

$$r_{DIST-NET\_BASE\_AC} = \frac{R_{DIST-NET}}{Z_{BASE\_AC\_HV}} \quad (9)$$

$$x_{DIST-NET\_BASE\_AC} = \frac{X_{DIST-NET}}{Z_{BASE\_AC\_HV}} \quad (10)$$

3.2 Distribution transformer (optional)

This transformer is used when *traction transformer's* voltage differs from the distribution network's one. The equivalent reactance and resistance is usually specified in unitary transformer values ($r_{DIST-TRANSF}$ and $x_{DIST-TRANSF}$) and has to be converted to common unitary quantities ($r_{DIST-TRANSF\_BASE\_AC}$ and $x_{DIST-TRANSF\_BASE\_AC}$), eqns. (13) and (14). *Distribution transformer's* high voltage base impedance ($Z_{NOM\_AC\_HV\_DIST-TRANSF}$) and high voltage zone's common base impedance ($Z_{BASE\_AC\_HV}$) are use to execute the conversion.

$$Z_{NOM\_AC\_HV\_DIST-TRANSF} = 3 \cdot \frac{U_{NOM\_L-N\_AC\_HV\_DIST-TRANSF}^2}{P_{NOM\_AC\_3-PHASE\_DIST-TRANSF}} \quad (11)$$

$$Z_{BASE\_AC\_HV} = 3 \cdot \frac{U_{BASE\_L-N\_AC\_HV}^2}{P_{BASE\_AC\_3-PHASE}} \quad (12)$$

$$r_{DIST-TRANSF\_BASE\_AC} = r_{DIST-TRANSF} \cdot \frac{Z_{NOM\_AC\_HV\_DIST-TRANSF}}{Z_{BASE\_AC\_HV}} \quad (13)$$

$$x_{DIST-TRANSF\_BASE\_AC} = x_{DIST-TRANSF} \cdot \frac{Z_{NOM\_AC\_HV\_DIST-TRANSF}}{Z_{BASE\_AC\_HV}} \quad (14)$$

3.3 3-phase lines of connection between supply points and traction transformers

These lines connect *supply points* (or distribution transformers if they exist) with *traction transformers*. The reactance and resistance will be normally arranged in total values ($R_{3-PHASE\_LINES}$ and $X_{3-PHASE\_LINES}$) and will need to be converted to common unitary quantities ($r_{3-PHASE\_LINES}$ and $x_{3-PHASE\_LINES}$) using high voltage zone's base impedance ($Z_{BASE\_AC\_HV}$), eqn. (16) and (17).

$$Z_{BASE\_AC\_HV} = 3 \cdot \frac{U_{BASE\_L-N\_AC\_HV}^2}{P_{BASE\_AC\_3-PHASE}} \quad (15)$$

$$x_{3-PHASE-LINES\_BASE\_AC} = \frac{X_{3-PHASE-LINES}}{Z_{BASE\_AC\_HV}} \quad (16)$$

$$r_{3-PHASE-LINES\_BASE\_AC} = \frac{R_{3-PHASE-LINES}}{Z_{BASE\_AC\_HV}} \quad (17)$$



3.4 Traction transformers

Those elements provide medium voltage to the railways' electric power system. The equivalent reactance and resistance is known in unitary transformer values ($r_{TRAC-TRANSF}$ and $x_{TRAC-TRANSF}$) so they have to be passed to common unitary quantities of the AC zone ($r_{TRAC-TRANSF\_BASE\_AC}$ and $x_{TRAC-TRANSF\_BASE\_AC}$), eqns. (20) and (21). Traction transformer's high voltage base impedance ($Z_{NOM\_AC\_HV\_TRAC-TRANSF}$) and high voltage zone's common base impedance are use to carry out the adaptation.

$$Z_{NOM\_AC\_HV\_TRAC-TRANSF} = 3 \cdot \frac{U_{NOM\_L-N\_AC\_HV\_TRAC-TRANSF}^2}{P_{NOM\_AC\_3-PHASE\_TRAC-TRANSF}} \quad (18)$$

$$Z_{BASE\_AC\_HV} = 3 \cdot \frac{U_{BASE\_L-N\_AC\_HV}^2}{P_{BASE\_AC\_3-PHASE}} \quad (19)$$

$$x_{TRAC-TRANSF\_BASE\_AC} = x_{TRAC-TRANSF} \cdot \frac{Z_{NOM\_AC\_HV\_TRAC-TRANSF}}{Z_{BASE\_AC\_HV}} \quad (20)$$

$$r_{TRAC-TRANSF\_BASE\_AC} = r_{TRAC-TRANSF} \cdot \frac{Z_{NOM\_AC\_HV\_TRAC-TRANSF}}{Z_{BASE\_AC\_HV}} \quad (21)$$

3.5 DC network

It is composed by the railway's overhead contact line, which transports the electric energy and are purely resistive; and by the train, moving charges that demand electrical power. The power base quantity will be identical to the AC one, eqn. (1), and AC low voltage and DC voltage are associated through the relationship of the rectifier, eqn. (3). Resistance of DC section (R_{LINES}) of the circuit have to be transformed to DC unitary quantities ($r_{LINES\_BASE\_DC}$) using the base impedance of this part of the circuit (Z_{BASE_DC}), eqn. (23).

$$Z_{BASE\_DC} = \frac{U_{BASE\_DC}^2}{P_{BASE\_DC}} \quad (22)$$

$$r_{LINES\_BASE\_DC} = \frac{R_{LINES}}{Z_{BASE\_DC}} \quad (23)$$

3.6 Equivalent DC resistance

The objective of solving a single DC circuit requires using an equivalent DC resistance (R_{eq}) to obtain the real voltage and intensity current. In the AC zone, impedance consists in reactance, which causes voltage drop but not power lose; and in resistance, which causes voltage drop and also power losses. When a model is developed to emulate this electrical performance, it must origin both voltage drop (due to the reactance and resistance combination) and power losses (due to resistance). The method used to complete an expression that complies with the conditions given before will be exposed at the next point providing the equivalent resistance shown in eqn. (35).

The resultant circuit can be solved with a power flow, and because of DC features, angles will not have to be work out, simplifying the resolution.



3.7 Total quantities

After the circuit resolution, every transformation can be undone; so voltages, currents and power flows are expressed in their real values.

3.8 Power losses correction

The results achieved with this method are quite acute, but if a better precision is needed for power losses, a correction can be made.

The electrical analysis of this adjustment is done in the following point, being eqn. (47) the expression that may correct the power losses.

4 Electrical analysis

Figure 2 matches up to an electrical module and angle study. Voltage applied to the system at the *supply point* is V_i . The power transmission system is composed of transformers, 3-phase AC lines, and DC lines; and all together could be represented as an impedance (provided the quantities have a common base), designated Z as I is shown in eqn. (24). Impedance will have resistance (R) due to transformers and DC lines and reactance (X) because of transformers and AC lines, eqns. (25) and (26) respectively.

$$\bar{Z} = R + jX \quad (24)$$

$$R = R_{DIST-NET\_BASE\_AC} + R_{DIST-TRANSF\_BASE\_AC} + \quad (25)$$

$$+ R_{3-PHASE-LINES\_BASE\_AC} + R_{TRAC-TRANSF\_BASE\_AC} R_{LINES\_BASE\_DC}$$

$$X = X_{DIST-NET\_BASE\_AC} + X_{DIST-TRANSF\_BASE\_AC} + \quad (26)$$

$$+ X_{3-PHASE-LINES\_BASE\_AC} + X_{TRAC-TRANSF\_BASE\_AC}$$

Trains are exposed to voltage V_o demanding an intensity current I . V_o must have the same angle as I because the charge feature of them is purely resistive. On the other hand, V_i will have a different angle value produced by the reactance effect.

The method used simply tries to get the equivalent resistance value which makes V_o in the DC model to be as similar as possible as it is in the AC one. The principal hypothesis is to suppose that the angle between V_i and V_o is very small so the Taylor series can be used instead of trigonometric functions.

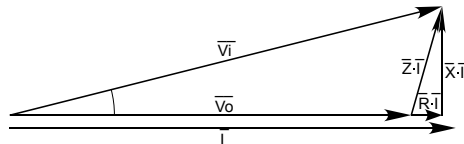


Figure 2: Phasorial diagram of relevant magnitudes in the real case.

Substituting Taylor series into trigonometric functions, eqns. (27), (28) and (29); an expression for $\cos(\varphi)$ can be accomplished, eqn. (30).

$$\sin(\varphi) = \frac{X \cdot I}{V_i} \quad (27)$$

$$\varphi \approx \frac{X \cdot I}{V_i} \quad (28)$$

$$\cos(\varphi) \approx 1 - \frac{1}{2} \cdot \varphi^2 \quad (29)$$

$$\cos(\varphi) \approx 1 - \frac{1}{2} \cdot \left(\frac{X \cdot I}{V_i} \right)^2 \quad (30)$$

Making use of the previous approximation, eqn. (30), the equivalent resistance, eqn. (34), will be calculated as the division of the voltage drop by the current intensity.

$$\Delta V = V_i - V_o \quad (31)$$

$$V_o = V_i \cdot \cos(\varphi) - R \cdot I \approx V_i - \frac{1}{2} \cdot \frac{(X \cdot I)^2}{V_i} - R \cdot I \quad (32)$$

$$\Delta V \approx \frac{1}{2} \cdot \frac{(X \cdot I)^2}{V_i} + R \cdot I \quad (33)$$

$$R_{eq} = \frac{\Delta V}{I} \approx \frac{1}{2} \cdot X^2 \frac{I}{V_i} + R \quad (34)$$

Finally, the approximation is settled as a function of the nominal demanded power through the respective element and the input voltage, eqn. (35).

$$R_{eq} \approx \frac{1}{2} \cdot X^2 \frac{P}{V_i^2} + R \quad (35)$$

The result after the conversion in DC magnitudes is shown in Figure 3.

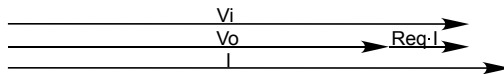


Figure 3: Phasorial diagram of relevant magnitudes in the simplified case.

This result depends on the demand of power, but the estimation has to be standardized to be applied in a general way for every component in the AC part. A power value must be introduced into the equation and it should be the real one demanded by the trains from this supply point, but it also should be constant. Because of that, a study to know which electric power grants better approximations needs to be done.

Below the real performance and the model are compared to identify the errors made because of approximations. Nominal power introduced in the equivalent resistance eqn. (35), and demanded power are the parameters that will determine the magnitude of errors. An example is proposed to establish comparisons between both models. It will receive typical values of input voltage, eqn. (36); resistance, eqn. (37); reactance, eqn. (38); and a grid of nominal power, eqn. (39), and demanded power, eqn. (40).



$$V_i = 1 pu \tag{36}$$

$$R = 0.05 pu \tag{37}$$

$$X = 0.15 pu \tag{38}$$

$$P_{NOM} = [0.25, 0.50, 0.75, 1.00, 1.25, 1.50] pu \tag{39}$$

$$P_{DEM} = [0.25, 0.50, 0.75, 1.00, 1.25, 1.50] pu \tag{40}$$

Equation (41) is used to solve the real case, Figure 2. Equation (42) will be also employed with equation (42) to explain the simplified model’s outcome, Figure 3. In both cases power consumption, eqn. (43), will be known and its value will affect the conclusions clearly.

$$V_i^2 = (V_o + R \cdot I)^2 + (X \cdot I)^2 \tag{41}$$

$$V_i = V_o + R_{eq} \cdot I \tag{42}$$

$$P_{DEM} = V_o \cdot I \tag{43}$$

Once both models have been worked out for the variation of parameters nominal power and demanded power, eqn. (39) and (40); voltage output, current and power losses are contrasted.

Relative errors (ε) are obtained by means of eqn. (44) and applied to variables V_o , I and L (power losses).

$$\varepsilon_{P_{NOM}, P_{DEM}} = \frac{\text{abs}\left(\text{Value}_{P_{NOM}, P_{DEM}} \Big|_{REAL} - \text{Value}_{P_{NOM}, P_{DEM}} \Big|_{MODEL} \right)}{\text{Value}_{P_{NOM}, P_{DEM}} \Big|_{REAL}} \tag{44}$$

Mean errors ($\bar{\varepsilon}$) for each nominal power value is also useful to rate the model’s goodness of fit. It is achieved using the probability each power consumption value has, eqn. (46).

$$Prob_{POWER} = [0.30, 0.20, 0.30, 0.15, 0.10, 0.05] \tag{45}$$

$$\bar{\varepsilon}_{P_{NOM}, P_{DEM}} = \sum_{i=P_{DEM}} \varepsilon_{P_{NOM}, i} \cdot Prob_i \tag{46}$$

Table 1 illustrates relative errors between real solution and model’s solution for V_o module. Table 2 shows mean errors for each nominal power value.

Table 1: Relative error between real and model’s V_o module value.

		P _{NOM} [pu]						
		0.25	0.50	0.75	1.00	1.25	1.50	0.25
P _{DEM} [pu]	0.25	0.0020917	0.074215	0.15064	0.22719	0.30386	0.38065	0.0020917
	0.50	0.18798	0.020522	0.14754	0.3162	0.48547	0.65536	0.18798
	0.75	0.6485	0.36887	0.087518	0.1956	0.48052	0.76727	0.6485
	1.00	1.5576	1.1343	0.70685	0.27516	0.16092	0.60172	1.5576
	1.25	3.3017	2.6832	2.0554	1.418	0.77029	0.11194	3.3017
	1.50	7.0122	6.101	5.1686	4.2134	3.2336	2.2273	7.0122

Table 2: Mean error between real and model’s V_o module value.

P _{NOM} [pu]	0.25	0.50	0.75	1.00	1.25	1.50
$\bar{\varepsilon}$ [%]	1.1472	0.88054	0.67096	0.58382	0.59525	0.68826

Table 3 illustrates relative errors between real solution and model’s solution for *I* module. Table 4 shows mean errors for each nominal power value.

Table 3: Relative error between real and model’s *I* module value.

		P _{NOM} [pu]						
		0.25	0.50	0.75	1.00	1.25	1.50	0.25
P _{DEM} [pu]	0.25	0.0020916	0.07427	0.15087	0.22771	0.30478	0.3821	0.0020916
	0.50	0.18763	0.020518	0.14775	0.3172	0.48784	0.65969	0.18763
	0.75	0.64432	0.36752	0.087441	0.19599	0.48284	0.7732	0.64432
	1.00	1.5337	1.1215	0.70188	0.2744	0.16118	0.60536	1.5337
	1.25	3.1961	2.6131	2.014	1.3981	0.76441	0.11182	3.1961
	1.50	6.5527	5.7502	4.9146	4.043	3.1323	2.1787	6.5527

Table 4: Mean error between real and model’s *I* module value.

P _{NOM} [pu]	0.25	0.50	0.75	1.00	1.25	1.50
$\bar{\mathcal{E}}$ [%]	1.1087	0.85369	0.65346	0.57367	0.59109	0.68945

Table 5 illustrates relative errors between real solution and model’s solution for *Power losses*. Table 6 shows mean errors for each nominal power value.

Table 5: Relative error between real and model’s *Power losses* value.

		P _{NOM} [pu]						
		0.25	0.50	0.75	1.00	1.25	1.50	0.25
P _{DEM} [pu]	0.25	2.8082	5.782	8.7649	11.757	14.759	17.77	2.8082
	0.50	2.427	5.5817	8.7582	11.957	15.178	18.422	2.427
	0.75	1.4918	4.85	8.2479	11.686	15.167	18.689	1.4918
	1.00	0.31731	3.2685	6.92	10.64	14.43	18.294	0.31731
	1.25	3.6545	0.177	4.1135	8.1609	12.325	16.614	3.6545
	1.50	10.22	6.1731	1.9591	2.4361	7.0288	11.838	10.22

Table 6: Mean error between real and model’s *Power losses* value.

P _{NOM} [pu]	0.25	0.50	0.75	1.00	1.25	1.50
$\bar{\mathcal{E}}$ [%]	2.6995	5.1225	8.4028	11.958	15.762	19.62

Because of errors in estimation of power losses is not as good as it would be desirable, a correction is applied to improve the adjustment, eqn. (47).

$$L_{CORRECTED\_MODEL} = L_{MODEL} - (R_{eq} - R) \cdot I^2$$

(47)

Table 7 illustrates relative errors between real solution and model’s solution for corrected *Power losses*. Table 8 shows mean error for each nominal power value.

Table 7: Relative error between real and corrected model’s *Power losses* value.

		P _{NOM} [pu]						
		0.25	0.50	0.75	1.00	1.25	1.50	0.25
P _{DEM} [pu]	0.25	0.0041832	0.14859	0.30196	0.45593	0.61049	0.76566	0.0041832
	0.50	0.37491	0.041036	0.29572	0.6354	0.97805	1.3237	0.37491
	0.75	1.2845	0.73375	0.17488	0.39226	0.96789	1.5522	1.2845
	1.00	3.0442	2.231	1.3995	0.54884	0.32163	1.2144	3.0442
	1.25	6.2901	5.1579	3.9875	2.7767	1.523	0.22351	6.2901
	1.50	12.676	11.17	9.5876	7.9226	6.1665	4.31	12.676

Table 8: Mean error between real and corrected model’s *Power losses* value.

P_{NOM} [pu]	0.25	0.50	0.75	1.00	1.25	1.50
$\bar{\epsilon}$ [%]	2.181	1.6818	1.2903	1.1377	1.178	1.3801

Information given in Table 1, Table 3, Table 5 and Table 7 is illustrated graphically in Figure 4.

5 Conclusion

This model is being used in real cases to calculate the electrical power supply for the Spanish railway’s power system. The achieved results, making use of the correct nominal power for each element, provide a fine approximation to the real solution.

Working with systems that combine AC and DC parts is facilitated by the exposed method, reaching a DC equivalent circuit that can be solved or studied simply.

As it is exposed in Figure 4, the best approximation is achieved when $P_{NOM}=1[pu]$.

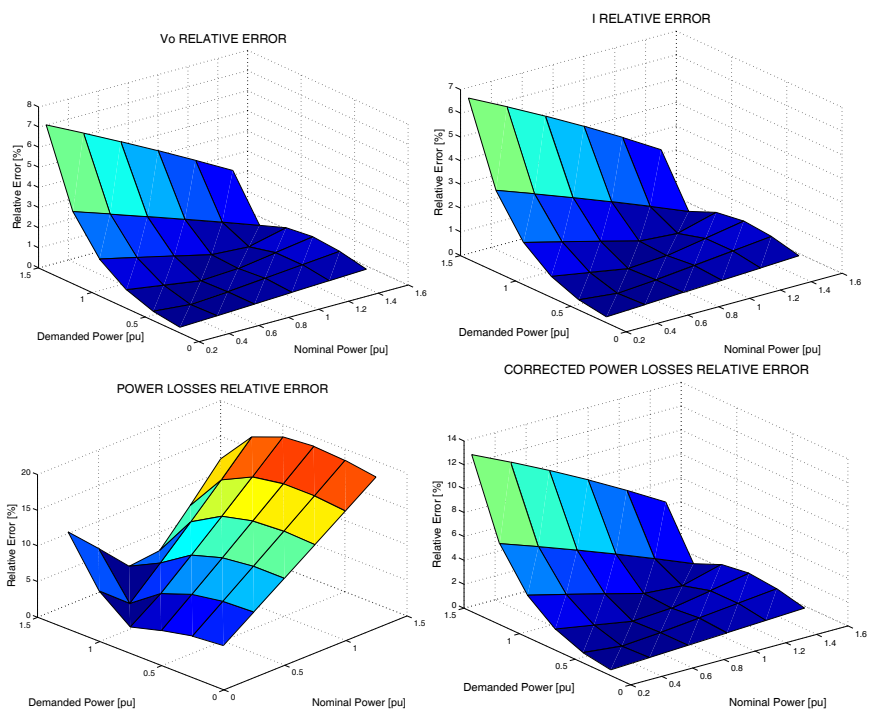


Figure 4: Relative error between real and corrected model’s variables.



References

- [1] Kundur, P., *Power system stability and control*, McGraw-Hill, pp. 468-497, 1993.
- [2] Kaller R., Allenbach J.M., *Traction électrique*, Presses polytechniques et universitaires romandes: Lausanne, 1995.
- [3] Profillidis V. A., *Railway Engineering*, Ashgate, 2000.
- [4] Hill R. J., Brillante S., Souza C. R. de et al., *Electrical material data for railway track transmission line parameter studies*, IEE Proceedings on Electrical Power Applications 146 (No 1, January 1999): 60-68.
- [5] Hill R. J., Carpenter D. C., *Determination of rail internal impedance for electric railway traction system simulation*, IEE Proceedings B 138 (No 6, November 1991): 311-321.



This page intentionally left blank

A user interface for the representation of the dynamic results on the pantograph-catenary interactions

D. Cebrián¹, T. Rojo², A. Alberto¹, E. Arias², F. Cuartero²
& J. Benet²

¹*Albacete Research Institute of Informatics,
University of Castilla-La Mancha, Spain*

²*Escuela Politécnica Superior de Albacete,
University of Castilla-La Mancha, Spain*

Abstract

In this paper a user-friendly environment for the representation of the dynamic results on the pantograph-catenary interactions is presented. This environment consists of an interactive interface for the representation of graphics and another for the representation of video. By means of the first interface for the representation of graphics, the user can interact with the different graphic representations obtained starting from the data generated by the dynamic calculation of catenaries. The objective of the second interface is to generate a video containing the simulation of the interaction between the pantograph and the catenary. The video sequence, generated by the user through its interactions, can be recorded in AVI format. These two interfaces have been integrated in a user-friendly, interactive and graphically oriented tool software called CALPE. This tool is used for the static and dynamic analysis of a catenaries system, which is shown by means of a real case study.

Keywords: pantograph-catenary interaction, user-friendly interface, graphical representation, video generation.

1 Introduction

Nowadays we are living in a period of growing economic globalization, a progressive removal of the frontiers of countries and a greater communication



between them. All these factors influence the means of transport that, for some time, have had to improve and adapt to the new situation, being quicker, resistant and efficient besides knowing how to respond in each moment to the necessities of society.

The railway transportation does not want to lag behind; therefore achieving quicker and more efficient trains has become one of the most important goals at the present time. For this reason an exhaustive study should be made on the operation of the railroad, seeing where the limitations of this means arise, and where we can obtain the most important enhancements.

One of the most important elements that it is necessary to keep in mind in the rail development is the electrical system, which provides to the locomotive with the driving force for their operation. The transmission of the electricity is carried out from the system formed by the group of wires placed throughout the railroad, named catenary, until the pantograph of the train. This transference of electrical energy provides the necessary force to the train to go throughout the railroad system. Therefore, the contact among those two elements should be as constant as possible, avoiding any take off that impedes the transmission of the electricity with the consequential loss of speed of the train and other associate problems. In this scenario, the system pantograph/catenary, with its dynamic behaviour, is a crucial component (see [1–6]) because at high speeds it is very difficult to guarantee the permanent contact of the head of the pantograph and the contact wire.

To obtain relevant benefits in the circulation of railway, the contact force among the pantograph and the catenary has to be maintained as even as it is possible, avoiding the lost of contact. This part is the most critical in the transmission of the electrical energy in modern high-speed trains [4], and therefore, many studies have been made in this field; regarding the design of new models of pantographs and catenaries, or about the simulation of this interaction [6], obtaining results used in the new installations with the objective of obtaining a permanent contact between the pantograph and the contact wire.

To work directly with these data and to understand them can be a very difficult and, sometimes, impossible work because of the great quantity of data. For this reason, two user-friendly graphic interfaces has been developed in which the data are represented in different ways and by means of which we can interact with the different data representations obtaining a better understanding of the data. The first interface will take charge of the representation of graphics, starting from the data generated by the dynamic calculation of catenaries, being able to select the graph type to view, to vary the range of the study, to select the element (wire or pantograph) to study, etc. By means of the second interface, the user will be able to interact in the reproduction of the video, to modify the reproduction speed, to change the scale of the wires, to change the way of the reproduction, to control the take off, etc.

The paper is structured as follows, in sections 2 and 3, both interfaces will be presented, in section 4, we show the integration in the software tool CALPE, which is currently used by the company Renfe in Spain, and we present the



application of this tool to a real case study. Finally, the conclusions and the future work will be commented in the last section.

2 Interface for the representation of graphical results

This interface generates a group of interactive graphics obtained from the data generated by means of the study of the interaction of the pantograph in movement and the catenaries system. This study is named dynamic calculation of catenaries [5]. To carry out this study, the pantograph/catenary system described in the previous section has been used.

By means of this interface we can interact with five types of graphics, being able to select in each moment the one in which we are interested. These types of graphics are explained in the following subsections.

2.1 Efforts in front of distance

This picture represents the force that makes the pantograph on the contact wire/s in each point of the journey. By means of the interface created to interact with this picture we can select the range to study. We can also select the pantograph that will be studied and to see the force that it makes on each contact wire or on several at the same time. By means of the button “*Restaurar gráfica*” the maximum studio range allowed by the data is selected. Lastly, the option “*Filtro*” allows the filtered data to be seen.

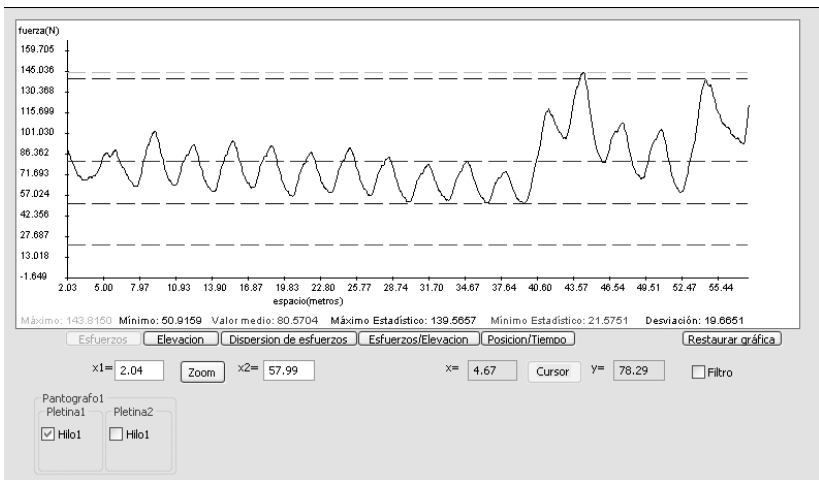


Figure 1: Interface to interact with efforts in front of distance picture.

2.2 Elevation in front of distance

In this picture the elevation of the contact wire in each point of the journey is represented. This elevation is produced by the force that exercises the pantograph



on the wire. Through the interface shown in figure 2, you can select the wire to study, besides adjusting the studio range like in the previous case.

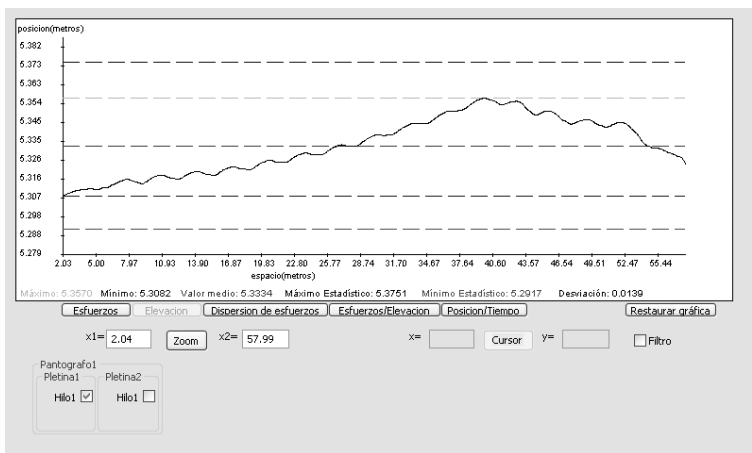


Figure 2: Interface to interact with elevation in front of distance picture.

2.3 Dispersion of efforts:

This picture shows the percentage of times that a certain force on the contact wire/s is repeated. Each reading will rarely be repeated so the forces are grouped in ranges that go from 0.1 to 0.9 N. This range can be modified by means of the slider “Tolerancia”. Also, like in the previous cases, you can also select the pantograph and the wires to study. In this case, the studio range cannot be modified because it is fixed by the previous graphics, but a zoom on the image can be realized.

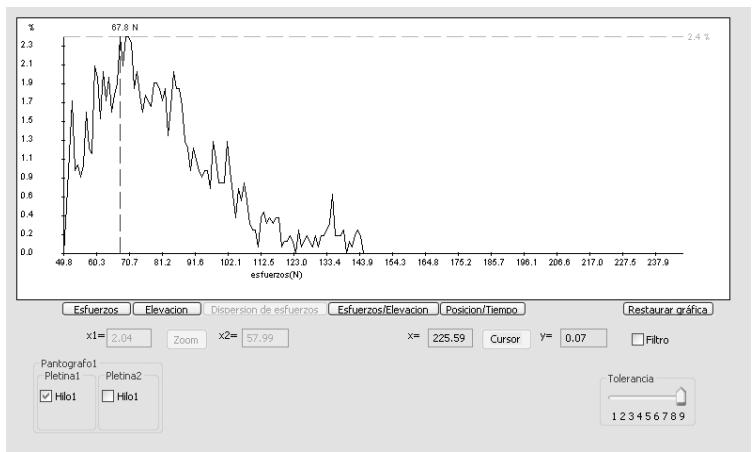


Figure 3: Interface to interact with dispersion of efforts picture.



2.4 Efforts/elevation in front of distance:

The first two graphics are superimposed using two vertical scales. The interface is also a mixture of the first two, allowing the height of a wire to be displayed, and several wires to see the efforts that are carried out on them. Here, the study range can be modified.

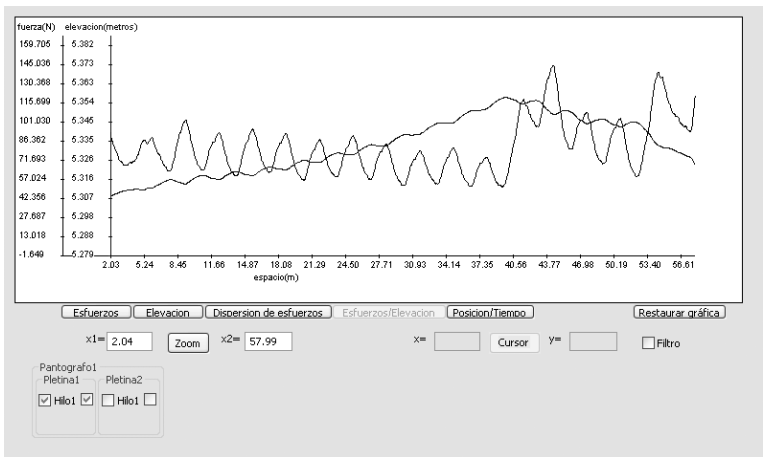


Figure 4: Interface to interact with efforts/elevation in front of distance picture.

2.5 Position in front of time

This shows the position (vertical) of a fixed point (g or q) throughout the time that the journey lasts. By means of the interface you can select the point to be observed. As in the dispersion of efforts graph the study range can not be modified and a zoom on the image can be realized.

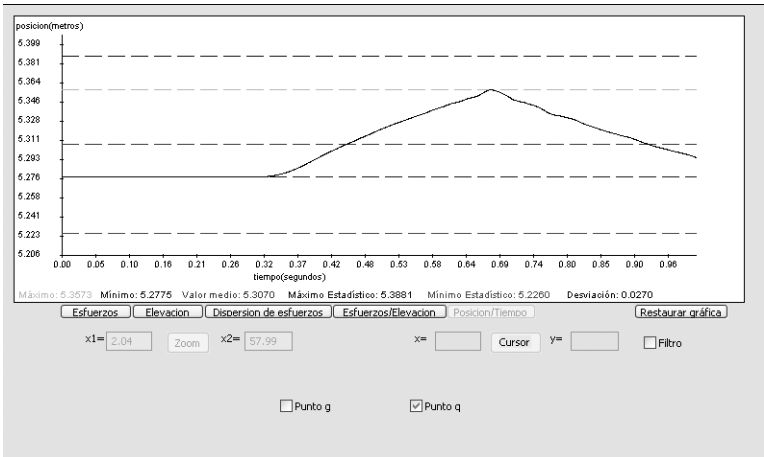


Figure 5: Interface to interact with position in front of time picture.



These graphics will be saved in jpeg format. As the component is interactive, the graphics will be able to be modified during the run of the component; therefore, two versions of the graph will be saved: the initial (unmodified) and the modified one.

This component generates quite a lot of output files. Besides the jpeg files that contain the graphics, it also generates a .inf file for each jpeg file that contains information about the graph. The .inf files contain information about the appropriate graph.

Bearing in mind that there is a graph of each type for each pantograph and that for each graph there are two jpeg files (one for the original graph and another for the modified one) so the quantity of files that are generated is considerable. Another file generated is an .act file that indicates the name of the jpeg file that contains the graph that is currently showing.

3 Interface for the video generation

In the previous section, a great variety of graphics have been represented. In these graphics one could observe the force that carries out the pantograph to the contact wire according to their speed or of their position, the elevation of the contact wire, etc.

By means of this interface the obtained data with the dynamic calculation of catenaries [5] (using the pantograph/catenary system described in the introduction section has been used) will be used even more, because in this interface, a video will be generated and in this video, the forces carried out by the pantograph on the catenaries system will be able to be appreciated. In this video a catenaries system will be shown in two dimensions, formed by one span that will be covered by one or several pantographs, this will depend on the number of pantographs that the train had. While the pantographs cover the span, one will be able to observe the movement of the contact wire (in the event of that there are two contact wires one will be able to choose which is wanted to see), of the holder wire and of the droppers, and also the take off that take place between the contact wire and the pantograph will be able to be seen.

In figure 6 a catenaries system is shown; this system is being covered by a pantograph. One can observe while the pantograph goes covering the span, some blue arrows appear, some up and others down. When an up arrow appears it means that the contact force between the pantograph and the contact wire is smaller than the showed in the text box "*Fuerza de despegue*", and this force can be modified by the user. When the force is bigger than this value a blue down arrow appears. During the time that the contact force is smaller than the value indicated in the text box "*Fuerza de despegue*" the pantograph is blue. When the contact force is negative, this means that a real take off takes place between the contact wire and the pantograph, the arrows become red and the pantograph too. The number that appears under the pantograph indicates the value of this contact force.

Besides being able to see this sequence of images so that all the data obtained in the calculation and the graphics generated can be understood, the user will be



able to interact with the video shown. By means of the slider placed to the left of the sequence the user will be able to vary the scales in run-time to see with more clarity the contact wire and the holder wire movement. Also the user will be able to vary the speed of the video playback by means of the slider placed in the inferior right part and, by means of the controls located under the sequence, the user will be able to carry out the typical actions of starting the playback forward, back, pause the playback and stop it. Another option that will allow this component is to record, in avi format, the video generated and to choose the compression method for the recording. The compression methods that the application will allow will vary in function of the video compression codes that the system has installed.

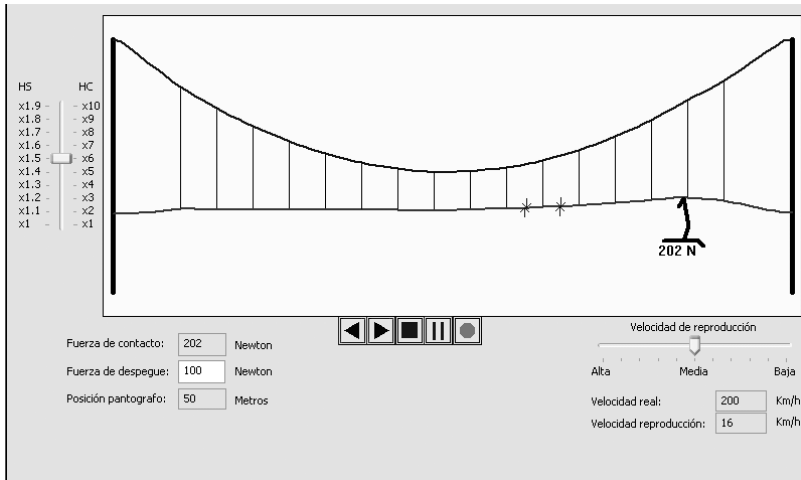


Figure 6: Interface for the video representation.

4 Integration in the CALPE application, and application to a case study

The interfaces presented in the previous sections have been implemented as part of a user-friendly, interactive and graphically oriented software tool, called CALPE. The software package has been developed on an object-oriented database system with a visual interface under Windows. This framework is supported in the Visual FoxPro environment, and it is currently used by RENFE, the Spanish company of railways, in the development of its electrical catenary systems.

In order to be able to be integrated in the application, these interfaces have been encapsulated in two ActiveX components. In figure 8 an image of this application is shown. By means of this tool, we can realize studies about dynamic calculation of catenaries or about static ones [7, 8]. This tool shows the



majority of features presented in [9]. In [10] another friendly tool for pantograph design and testing is showed.

To show the graphics and the video generated by these interfaces we present a real study done with this tool. The input data is a pantograph/catenary system with the following features:

• Type of catenary system:	cr220 CIMYESCLE
• Number of spans:	1
• Number of pantographs:	1
• Type of pantographs:	SHUNK 200
• Velocity of train:	200 km/h
• Span length:	60 m
• Catenary system height:	1400 mm
• Contact wire height:	5300 mm
• Section of holder wire:	184 mm ²
• Mechanics tension of holder wire:	2550 kg
• Section of contact wire:	150 mm ²
• Mechanics tension of contact wire:	1875 kg
• Number of contact wires:	1
• Section of dropper wires:	25 mm ²
• Number of droppers:	16
• Initial curvature:	30 mm

With this characteristics, we obtain a very complete report suitable for industrial application. It contains 55 pictures in 59 pages, and can be seen at the address <http://www.info-ab.uclm.es/fmc/publications/2006/graphics220.pdf>.

5 Conclusions and future work

In this paper two user-friendly interfaces for the representation of the dynamic results on the pantograph-catenary interactions have been presented. The first one generates a group of interactive graphics starting from the data generated by means of the studio of the interaction of the pantograph in movement and the catenaries system and the second one generates a video which shows a catenaries system and the forces carried out by the pantograph on the catenaries system. Finally, how these interfaces have been integrated in the CALPE application has been explained.

Starting from the work presented in this paper, new necessities and new ideas appear in order to take advantage of the data obtained by means of the dynamic calculation of catenaries. Some of these ideas and necessities include the following.

- To expand the studio of the interaction pantograph-catenary to new kind of spans, what involves the generation of new graphics and videos for spans of different typology.



- To provide a bigger interactivity to the graphics and the video so that the user has a bigger freedom of work.
- To represent graphics in 3-dimensions to be able to observe the vertical oscillations of wires like the horizontal ones.
- In the representation of the video, to allow the simulation of more than a span so that, one can observe the behaviour of the spans in those which the pantograph is not.
- Finally, to carry out a simulation in 3-dimensions of the whole catenaries system and be able to see the behaviour that will have in the reality each one of the elements.

References

- [1] Poetsch G., Evans J., Meisinger R., Kortum W., Baldauf M., Veitl A., Wallaschek J. "Pantograph/catenary dynamics and control", *Vehicle System Dynamics*, 28:159-195, 1997.
- [2] Poetsch G. and Wallaschek J. "Symulating the dynamic behaviour of electrical lines for high-speed trains on parallel computers", *International Symposium on Cable Dynamics*, Lige, 1993.
- [3] Simeon B. and Arnold M. "The simulation of pantograph and catenary: a PDAE approach", *Technical Report 1990*, Fachbereich Mathematik Technische Universitat Darmstadt, 1998.
- [4] Simeon B. and Arnold M. "Coupling DAEs and PDEs for Simulating the interaction of pantograph and catenary", *Mathematical and Computer Modelling of Dynamical Systems*, 6-2:129-144, 2000.
- [5] Arias, E., Benet, J., Bruis, P., Cuartero, F., Rojo, T. "Solution of the Dynamical Problem of Pantograph-Catenary Interaction: A High Performance Approach". *Proc. of the 7th International Conference on Computational Structures Technology*. Lisboa, Septiembre, 2004.
- [6] Larsson, T., Drugge, L., "Dynamic Behaviour of Pantographs due to different Wear Situations". *Computers in Railways VI*, WIT Press, Southampton, UK, pp. 869-880. ISBN 1-85312-598-9, (1998).
- [7] Benet, J., Cuartero, F., Rojo, T. "A Tool to calculate catenaries in railways". *Computer in Railways VII*. WIT Press, Southampton, UK. pp. 535-544, 2000.
- [8] Benet, J., Montesinos, J., Cuartero, F., Rojo, T., Arias, E. "Advanced algorithm to calculate mechanical forces on a catenary". *Computer in Railways IX*. WIT Press, Southampton, UK. pp. 857-866. Mayo, 2004.
- [9] Drugge, L., Larsson, T., Stensson, A. "Modelling and simulation of catenary-pantograph interaction". *Vehicle System Dynamics*, 33. pp. 490-501. 1999.
- [10] Balestrino, A., Bruno, O., Landi, A., Sani, L. "PANDA: a friendly CAD tool for Pantograph design and testing". *Computer in Railways IX*. WIT Press, Southampton, UK. pp. 817-826. Mayo, 2004.



This page intentionally left blank

Section 15

Dynamics and wheel/rail interface

This page intentionally left blank

Dynamic identification of a 1:5 scaled railway bogie on roller rig

N. Bosso, A. Gugliotta & A. Somà

Politecnico di Torino, Dipartimento di Meccanica, Italy

Abstract

A Scaled Roller Rig is an experimental device to carry out laboratory tests on railway vehicles. The advantages of this method consist of low costs, low testing time and good repeatability. Different testing conditions, such as different friction, wear of profiles and axle load can be easily reproduced. Due to this reason many authors use roller rig to study railway dynamic, wear, noise and other railway problems. This work describes the experimental tests carried out to analyze the dynamical behaviour of a 1:5 scale railway bogie on a roller rig in tangent track conditions. The Roller Rig has been designed and realized at Politecnico di Torino and the first tests have been conducted on a single suspended wheelset.

Tests performed on a single wheelset are used to validate the wheel/rail and wheel/roller contact algorithm and to adjust the parameters of numerical models. On the other hand the study of a complete railway bogie allows one to simulate the behaviour of a real vehicle and can be used in the design phase of new vehicles. For instance, it is possible to evaluate the effect of changes brought to the suspension system on the running dynamics of the vehicle.

Keywords: wheel-rail, dynamics, contact, roller-rig.

1 Introduction

The use of Roller-Rig allows one to simulate the behaviour of real vehicles under well known testing conditions. If the tests are performed on reduced scale prototypes, it is possible to significantly reduce the costs.

Unfortunately the railway vehicle has a different behaviour on the track respect on Roller Rig, this is mainly due to finite curvature of the rollers, which modify the shape of the contact patch and add an instable gravitational effect



during the yaw rotation of the wheelset. This primary effect leads to a higher instability of the wheelset on roller rig, depending on the roller radius (higher radius leads higher instability); this aspect can be analytically determined by modifying the equations governing the kinematical creepages [6, 8, 10].

Because the shape of the contact area change due to the finite curvature of the rollers, it is possible to modify the transversal curvature of the roller in order to achieve the same shape. For example on figure 1, the elliptical contact area are shown for wheel-rail and wheel-roller contact, considering a conical wheel (0.3 rad) and a ratio $R_{\text{wheels}}/R_{\text{Roller}} = 0.5$. The two shapes are different, but if the transversal curvature radius of the roller is modified (0.025 m instead of 0.06 m, in 1:5 scale), it is possible to obtain the same a/b ratio again.

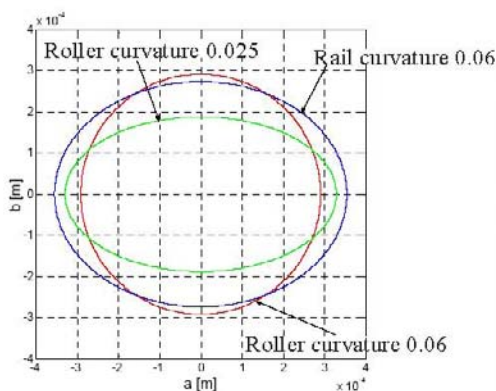


Figure 1: Different contact shape considering a conical (0.3 rad) wheel over a rail (transversal curvature radius 0.06 m) or a roller (0.06 m and 0.025 m).

Another effect, which differentiates the dynamic on roller rig respect to the dynamic on rails, is the fact that the contact point moves forward during yaw movement of the wheelset more on rollers than on rails (contact point shift [8]). This is negligible on straight track but becomes important during curving. For this reason, to simulate the curve behaviour on roller rig, it is required to supply to the rollers a number of controlled movements in order to keep each wheel on the top of the corresponding roller. Furthermore, on narrow curve, if the contact point shift produces a flange contact located forward respect the tread contact, this cannot be reproduced on a roller rig.

The roller rig built at Politecnico di Torino is used to perform dynamic simulations on straight track using a single wheelset and an entire bogie (2 axles). The prototypes are realized in reduced scale (1:4 – 1:5) and therefore it is required to use a similitude approach in order to convert the results obtained on the scaled prototype to the real vehicle.

The choice of the appropriate similitude law is not unique, and in general it is not possible to reproduce all the physical laws on a scaled model at the same time. In fact the scaling factors which relates the scaled quantity to the unscaled



ones represents a linear relation and not all the phenomena involved are linear. In order to choose the most suitable law, different models proposed in literature have been analysed and compared [4]. The prototype has been designed according the similitude law proposed by Jaschinski [2], whose most important scaling factors are shown on table 1.

Table 1: Scaling factors according to Jaschinski's similitude (1:5).

Unit	symbol	Scaling factor
Length	φ_l	5
Time	φ_t	$\sqrt{5}$
Velocity	φ_v	$\sqrt{5}$
Acceleration	φ_a	1
Mass	φ_m	125
Inertia moments	φ_I	3125
Force	φ_F	125
Creep Forces	φ_T	125
Torque	φ_C	625
Friction Coefficient	φ_μ	1

The roller-rig has been realized using a modular approach, which allows one to simulate different types of vehicles; it is possible to modify the wheelset spacing (1.8–3.5 m in real scale), the axle load, the profiles conicity and the suspension characteristics (see fig. 2).

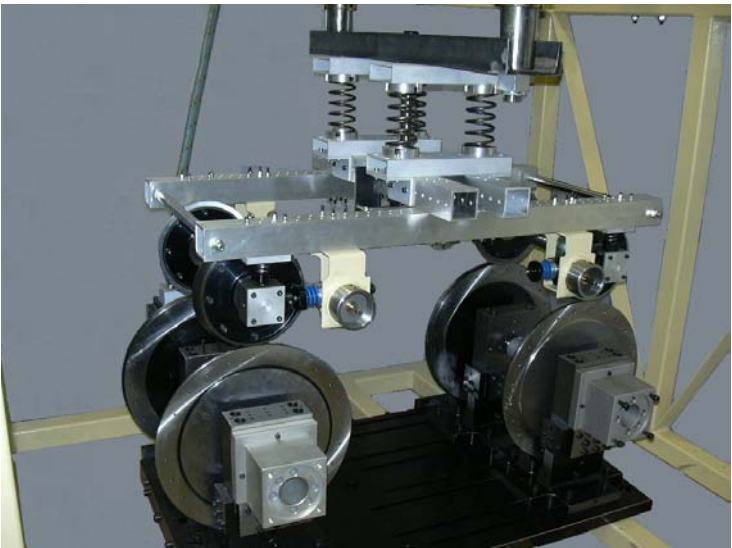


Figure 2: Roller Rig realized at Politecnico di Torino with a modular bogie.



2 Experimental device

The Roller Rig can be used to test a bogie prototype (fig. 2); in this case each pair of rollers is used to support a wheelset of the bogie. The rollers are independent and can be moved in the longitudinal direction in order to fit with the spacing of the bogie. The bogie itself has been designed with variable wheelset spacing in the range 1.8–3.5 m in real scale, in order to simulate every kind of railway vehicle (freight – high speed).

The primary suspension system has been studied with separate stiffness and preload regulation for each direction (lateral, longitudinal and vertical), the stiffness is provided by helical springs which can be easily replaced.

The axle load can be modified with additional loads applied on the load beam which is located over the secondary suspension stage and can move only in the vertical direction. The longitudinal motion of the bogie respect to the rollers is constrained by a pair of anti-symmetric traction bars connecting the centre of the bogie with the external structure.

The roller rig has been used to carry on tests on a single wheelset (fig. 3), in this case only a pair of rollers is used. The wheelset is suspended using a support structure providing separate stiffness in longitudinal, vertical and lateral direction.

The preload can be regulated for each direction using a screw device which integrates an estensimetric system to measure the load.

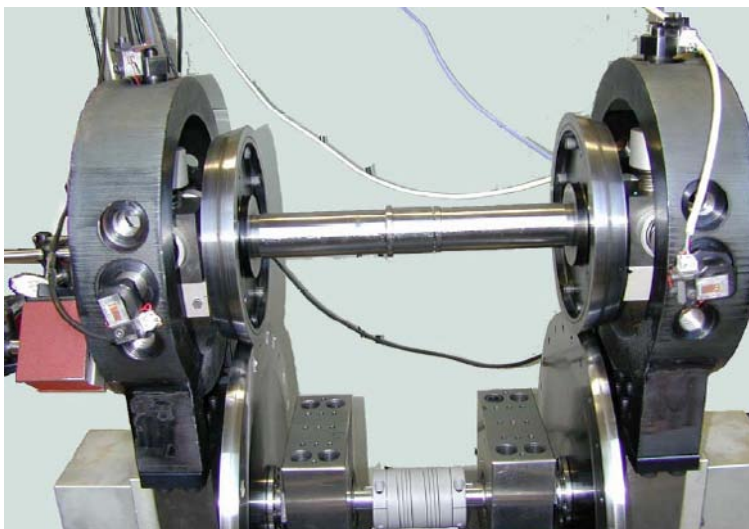


Figure 3: Testing device for a single wheelset.

2.1 Measuring systems and experimental identification

The measuring system is composed by accelerometers applied on the wheelset axle-box in longitudinal, lateral and vertical direction (fig. 4). In order to



measure the wheelset displacement capacitive transducers are used in lateral and longitudinal direction (LVDT X and Y on fig. 4).

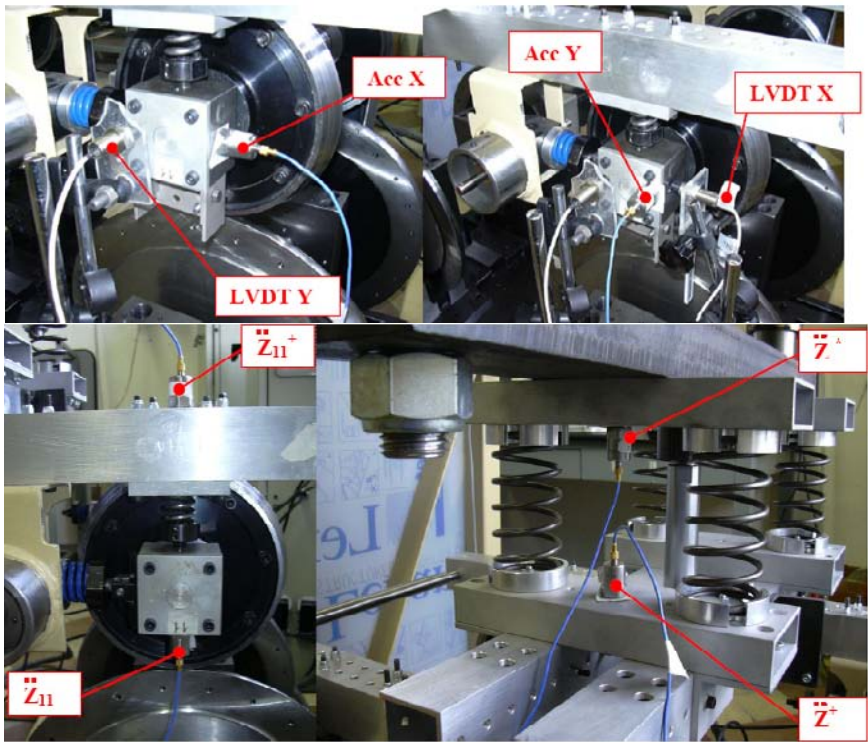


Figure 4: Accelerometers and displacement transducers.

Additional accelerometers have been mounted on the bogie frame and on the load beam in the vertical direction in order to measure the roller-coach transfer function. The forces acting on the suspension has been measured on the single suspended wheelset (fig. 1) while on the bogie an indirect measure has been done using extensimetric bridges applied on the bogie frame structure.

While in the case of the single wheelset tests, the “vehicle” characteristics can be easily identified (it simply consists on the inertial data and the shear and axial stiffness of the springs), in the case of the test on the bogie, the experimental identification is more complex.

The vehicle configuration used for the tests here described refers to a passenger high speed vehicle, with wheelset spacing equal to 2.7 m, high values of the longitudinal ($C_x = 5 \text{ MN/m}$) and lateral ($C_y = 10 \text{ MN/m}$) stiffness. The prototype required an accurate identification of all the parameters, performed via a Modal Analysis on the bogie suspended (fig. 5 left) and mounted on the rollers. This analysis has been preceded by static and dynamic (fig. 5 right) identification of each spring and of mass and inertia of each component.



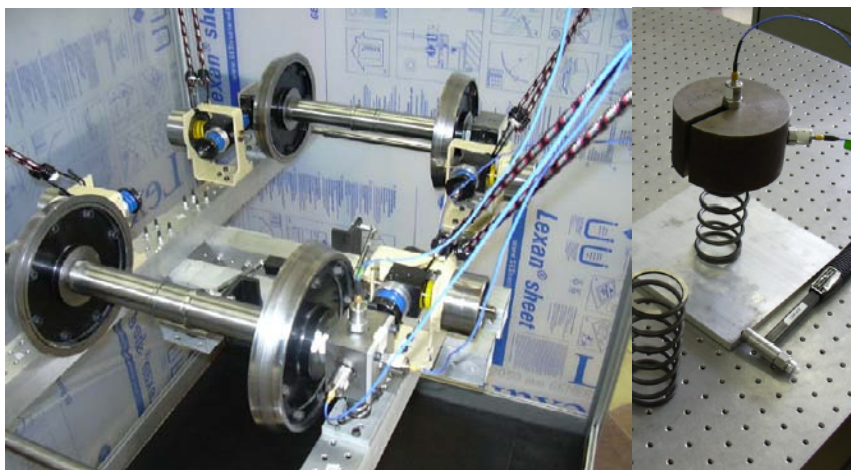


Figure 5: Modal analysis on the suspended bogie (left). Test on components (right).

3 Numerical model

In case of the single wheelset the numerical model is trivial regarding the vehicle description: the vehicle model is made of a single rigid body (the wheelset) constrained to the inertial system by the set of springs and to the rollers by the wheel-roller contact. This allows one to focus the attention on the contact problem.

The model of the single wheelset has been realized using a Multibody code (MSC/Adams see [12, 13]) where the wheel-roller contact has been described by a formulation developed by the authors [10, 11].

To simulate the behaviour of the boogie, a model of an entire vehicle has been realized using a commercial Multibody code (Simpack). In this way it is possible to validate the vehicle model using the Roller-Rig and subsequently to use the model to predict the behaviour of the vehicle in situation which cannot be simulated in the roller-rig, such as curving.

Of course this choice does not allow one to obtain the same modelling flexibility of the first approach (the contact model can't be modified) and the results are expected to be less accurate. This is not a big limitation since in this case the interest is to perform optimization on the vehicle characteristics and only relative indications are necessary: e.g. it is useful to know if increasing the stiffness the critical speed increase or not, but it is not important to know the exact value of the critical speed. However a detailed model has been developed (fig. 6) using the data obtained by the experimental identification.

The models have been used in order to reproduce the same conditions of the experimental simulation performed on the roller-rig. With this aim a straight track has been used, but on the ideal track the irregularities detected on the rollers have been superimposed repeating N-times the envelope length of the rollers.



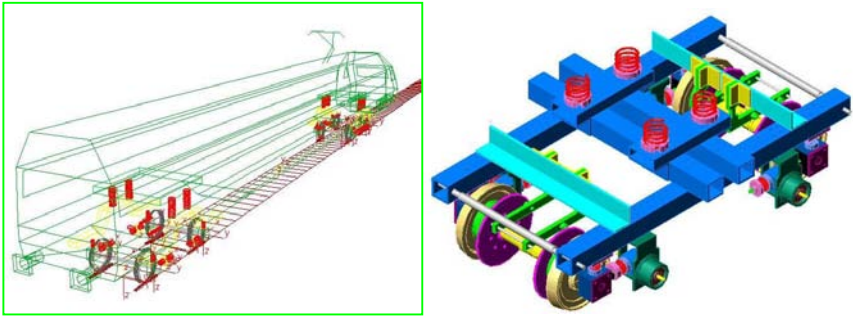


Figure 6: Numerical multibody model of the entire vehicle (left), detail of the bogie (right).

4 Results

Simulations have been performed in order to evaluate the stability of the vehicle: the rollers velocity has been increased since the critical speed has been reached.

Using the single suspended wheelset, it has been possible to determine the limit cycle diagrams (velocity-amplitude of oscillation) for different configuration (load/ conicity/ stiffness, see [13]).

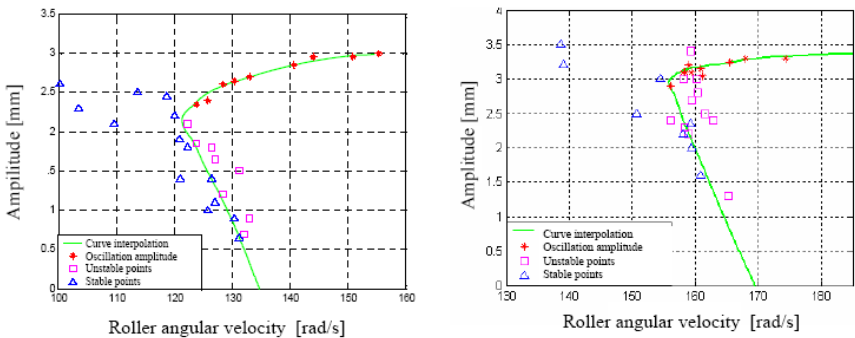


Figure 7: Limit cycles of instability obtained with conicity 0.3 (left) and 0.05 rad (right).

In this case the haunting motion of the wheelset has been excited with lateral impulsive forces of different amplitude. If the resulting oscillations are damped, the motion is stable; otherwise a fixed amplitude oscillation occurs: the amplitude value defines the upper limit cycle. The lower curve of the limit cycle is defined by those amplitude values which delimit the stable by the unstable region of the velocity-amplitude diagram.

Figure 7 shows the experimental limit cycles obtained with 0.3 and 0.05 rad conical profiles. It is evident that in reality it is possible to have instability even at lower velocity than the critical speed (intersection of the diagram with the x



axis). A new limit value is defined: the limit speed; between the limit speed and the critical speed instability occur only if the excitation is over a certain amplitude (lower limit cycle). The experimental values have been compared with the numerical model, and results are shown on table 2. Values of the critical speed and of the limit speed can be predicted with a very good approximation. In order to obtain an accurate value for the amplitude of oscillation, it is necessary to use a non-linear model for the springs in the shear direction; in fact large oscillations occur and use of linear stiffness is not sufficient.

Table 2: Numerical/experimental comparison conicity 0.3.

	Experimental	Numerical	% Difference
Linear stiffness model			
Critical speed [m/s]	23.2	23.3	0.8%
Limit speed [m/s]	21.1	21.8	3.3%
Oscillation amplitude [mm]	2.8	3.2	12.5%
Non linear stiffness model			
Critical speed [m/s]	23.2	23.3	1.5%
Limit speed [m/s]	21.1	22.2	5.6%
Oscillation amplitude [mm]	2.8	2.8	0.0%

4.1 Test performed on the bogie

Simulation performed on the bogie have are synthesized on table 3 using a high conicity wheelset (to simulate worn profiles) and different axle loads; results are indicated in 1:5 scale. In this case, due to the high stiffness, the limit cycles are “compacted”: the distance between limit and critical velocity is small.

Table 3: Experimental results with different axle load (conicity 0.3).

Velocity	Axle load	Acceleration		Displacement		Frequency
[m/s]	[Kg]	[m/s ²]		[mm]		[Hz]
		X	Y	X	Y	
40	42	0.069	6.32	0.6	5.5	5.5
35	47	0.501	2.73	0.32	1.74	6.3
30	52	0.107	4.66	0.9	3.9	5.2
27.5	52	0.117	5.01	1.1	4.7	5.5

Therefore the study has been performed considering the unstable behaviour. The axle load evidence a negative effect over the stability. This effect is complementary to the effect of wheel-rail friction. Unfortunately the friction coefficient change during the test, in particular, if instability occurs for long period, the friction coefficient increases. The friction coefficient has been measured only statically at the beginning of the test (value 0.38) and numerical simulations are performed using this value. During the test, when the friction increases higher amplitude oscillations can be observed and explains differences



respect the numerical model (table 4). Numerical Analysis shows the same behaviour as shown on figure 8: the instability is reached at almost the same velocity of the rollers (error lower than 10%). The experimental results show a more unstable behaviour (except at 35 m/s).

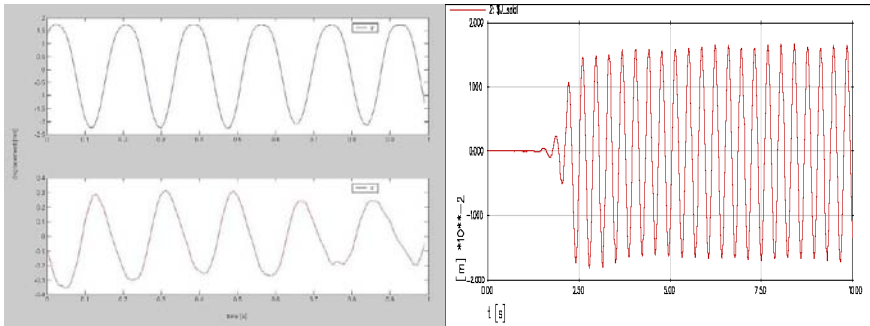


Figure 8: Experimental (left) and numerical behaviour (right) of the first wheelset at 40 m/s.

Large difference on the displacement and acceleration can be explained by the different friction coefficient and by the fact that suspension non-linearity has been neglected, and this could not be appropriate for large displacements.

Table 4: Numerical/experimental comparison (lateral direction).

Velocity	Axle load	Y Acceleration	Y Displacement	Frequency
[m/s]	[Kg]	[m/s ²]	[mm]	[Hz]
40	42	-30.4 %	-36.3 %	30.4 %
35	47	3.7 %	47.1 %	4.7 %
30	52	-35.6%	-31.3 %	20.7 %
27.5	52	-37.6 %	-30.6 %	18.2 %

5 Conclusions

The paper shows the activity performed on the Roller Rig realized at Politecnico di Torino, both on a single suspended wheelset than on a complete two-axle bogie. The tests performed on a single wheelset show a very good agreement with numerical results, and it is possible to evaluate not only the critical speed, but also the limit cycles of instability. The multibody numerical model allows an accurate identification of all the parameters and non-linearity. For this reason this system (single wheelset) can be adopted to perform investigations on the wheel-rail contact in order to improve the contact model. A high speed bogie is considered, and, due to the high stiffness in the primary suspension, the limit cycle vanishes. Amplitude of oscillation is strongly influenced by non-linearity (gap, friction) and in this case, due to the complexity of the model it is not easy to identify all the parameters with the same accuracy of the case of the single

wheel set. Anyway, results are useful in relative terms in order to perform optimisation on the suspension characteristics.

References

- [1] Iwnicki S.D., 1998, "Roller rig scaling", European Roller Rig Meeting, Manchester 27-8-1998.
- [2] Jaschinski A., 1990, "On the application of similarity laws to a scaled railway bogie model", DLR Institut für Dynamik der flugsysteme, Oberpfaffenhofen (Germany).
- [3] Iwnicki S.D., Wickens A. H., 1998, "Validation of a MATLAB Railway Vehicle Simulation Using a Scaled Roller Rig", Vehicle System Dynamics, pp. 257-270.
- [4] A. Jaschinski, H. Chollet, S. Iwnicky, A. Wickens and J. von Wurzen, "The application of roller rigs to railway vehicle dynamics", Vehicle System Dynamics, No. 31, pp. 345 – 392, 1999.
- [5] R.V. Dukkipati "A parametric study of the lateral stability of a rail bogie on a roller rig", Proc. Inst. Mech. Engrs. Part F, Vol. 213, pp. 39 -47, 1999.
- [6] N. Bosso, A. Gugliotta, E. Napoli, A. Somà, "Simulation of a scaled roller rig", 5th ADAMS/Rail Users' Conference; Haarlem, Netherlands – May 10th –11th, 2000.
- [7] N. Bosso, A. Gugliotta, A. Somà, "Progettazione di un banco prova in scala 1/5 per analisi sperimentale di carrelli ferroviari " - XXIX AIAS – Lucca, 2000.
- [8] N. Bosso, A. Gugliotta, L. Gusman, A. Somà. "Simulazione del comportamento dinamico di una sala ferroviaria per Roller-Rig" - XXIX AIAS – Lucca, 2000.
- [9] N. Bosso, A. Gugliotta, A. Somà, "Comparison of different scaling techniques for the dynamics of a bogie on roller rig"- 17th Symposium Dynamics of Vehicles on road and tracks - IAVSD 2001 - Technical University of Denmark. Copenhagen (Lyngby), August 20-24, 2001.
- [10] N. Bosso, A. Gugliotta, A. Somà, "Introduction of a wheel-rail and wheel-roller contact model for independent wheels in a Multibody code." – ASME/IEEE Joint Rail Conference, Washington, DC, 22-24 April 2001.
- [11] N. Bosso, A. Gugliotta, A. Somà, "Dynamic and design of a 1/5 scaled roller-rig: simulation of wheel rail contact". Rolling Contact Fatigue. Brescia, 2002.
- [12] N. Bosso, A. Gugliotta, A. Somà, "Experimental validation of a wheel-rail contact model using a scaled roller rig." 1st MSC-ADAMS Conference. London UK. 2002.
- [13] N. Bosso, A. Gugliotta, A. Somà, "Stability analysis of a single suspended wheelset on a 1/5 scale roller rig –comparison of experimental test versus multibody numerical simulations." PRORAIL 2003 – ZILINA (Slovak Republic).



On enhanced tilt strategies for tilting trains

B. Kufver¹ & R. Persson²

¹*Ferroplan, Sweden*

²*Bombardier, Sweden*

Abstract

Many railways have taken tilting trains into operation on lines with horizontal curves with small radii. Tilting trains have vehicle bodies that can roll inwards, thus reducing the lateral acceleration perceived by the passengers. Hence, tilting trains can run through curves at an enhanced speed. However, too much tilt can cause motion sickness among sensitive passengers. The tilt motions, generated by computer-controlled actuators should be optimised with care, taking the local track geometry and actual train speed into account. This paper presents tilt algorithms aimed at balancing conflicting objectives. Furthermore, the paper discusses the usefulness of route files (track geometry data) onboard the train and possible positioning systems.

Keywords: *tilting train, tilt algorithm, tilt strategy, alignment, cant, passenger comfort, motion sickness.*

1 Introduction

The idea of tilting trains on lines with curves with small radii was discussed already during the 1930s [1, 2]. Tilting trains have vehicle bodies that can roll inwards, thus reducing the lateral acceleration perceived by the passengers. Hence, tilting trains can run through curves at an enhanced speed. Nowadays, it is a mature technology that many railways have taken into revenue service.

However, experience shows that tilting trains can cause motion sickness among sensitive passengers [3–11]. The mechanisms behind motion sickness are not fully understood, but there is evidence that high roll velocities, generated by steep superelevation ramps in the tracks and by a high compensation ratio in the tilt system, contribute to the onset of motion sickness. Hence, too little tilting will cause discomfort caused by high lateral acceleration, while too much tilting may provoke motion sickness.



Most existing tilting trains use a fixed compensation ratio (i.e. the amount of tilt is proportional to the lateral acceleration). This paper presents an enhanced approach, where the compensation ratio depends on local track conditions and the train speed.

2 Vehicle motions on curves

2.1 Vertical curves

On circular curves in the vertical alignment, there will be an additional vertical acceleration which should be superimposed on the gravitational acceleration. The additional vertical acceleration depends on the vehicle speed to the power of two, and therefore, is higher for a tilting train running at an enhanced speed.

2.2 Circular curves in the horizontal alignment

When the vehicle is running on circular curves in the horizontal alignment, there will be a horizontal acceleration which depends on the vehicle speed to the power of two. By arranging cant and using a tilt system, the vehicle body will roll inwards at an angle $\mathcal{G}$. In the inclined vehicle body, the perceived lateral a_L and vertical accelerations a_V change according to eqns (1) and (2).

$$a_L = \frac{V^2}{R} \cdot \cos(\mathcal{G}) - \left(g + \frac{V^2}{R_v}\right) \cdot \sin(\mathcal{G}) \quad (1)$$

$$a_V = \frac{V^2}{R} \cdot \sin(\mathcal{G}) + \left(g + \frac{V^2}{R_v}\right) \cdot \cos(\mathcal{G}) \quad (2)$$

where V is vehicle speed in m/s, R is radius of horizontal curve in metres, $\mathcal{G}$ is total roll angle (depends on cant, suspension characteristics of the vehicle and tilt system), and R_v is radius of vertical curve in metres.

A reduction of a_L by an increased roll angle $\mathcal{G}$ is always correlated with increased roll motions and increased vertical acceleration a_V . (Perceived yaw and pitch motions are also affected by the roll angle, but these effects are not in focus in this paper.)

2.3 Transition curves in the horizontal alignment

Transition curves are track segments where horizontal curvature changes gradually between straight tracks and circular curves (and between circular curves of different radii). It is normal practice to arrange superelevation ramps at the same location as the transition curves. It is also normal practice to apply the tilt motions on the transition curves, with no intentional delays.



Motion quantities of particular interest are maximum values for lateral acceleration, lateral jerk (the rate of change of lateral acceleration), roll angle and roll velocity (the rate of change of roll angle). Figures 1–3 show different cases with respect to phase relations between the tilt motion and the track alignment.

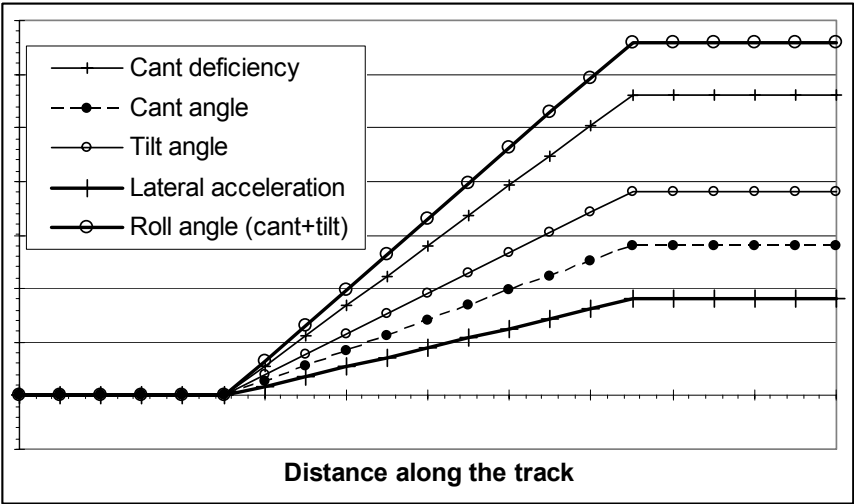


Figure 1: Entry transition: Motion quantities when the tilt motion is in phase with the alignment.

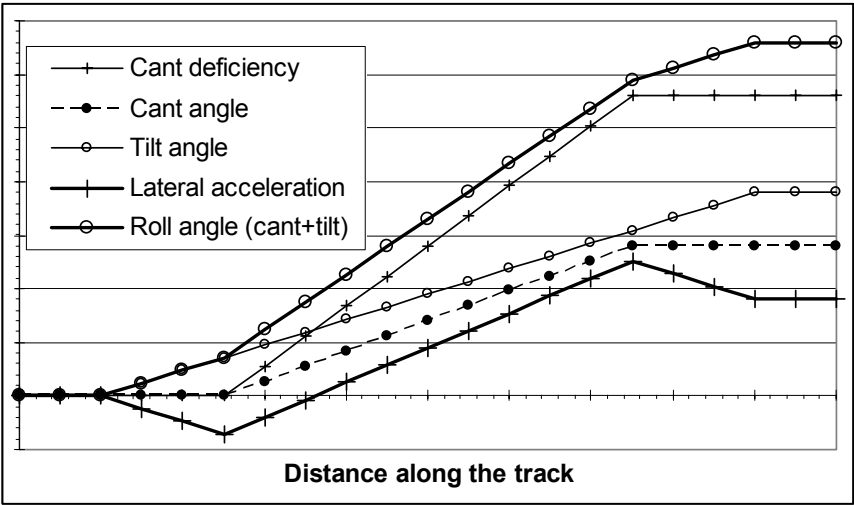


Figure 2: Entry transition: Motion quantities when the tilt motion is lengthened compared to the length of the transition curve.



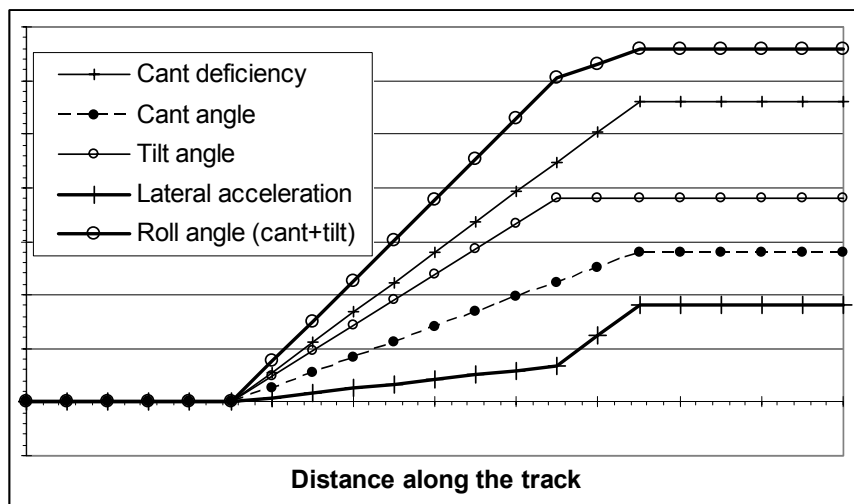


Figure 3: Entry transition: Motion quantities when the tilt angle saturates before the vehicle enters the circular curve.

In Figure 2, it can be seen that a longer tilt transition than the curve transition reduces the roll velocity (the rate of change of roll angle) but increases the maximum lateral jerk (the rate of change of lateral acceleration). If the mismatch between geometry and tilt occurs at the end of the entry transition, there will be a local maximum in the lateral acceleration, while if the mismatch occurs at the beginning of the entry transition there will be a change of direction of the perceived lateral acceleration. Current knowledge does not prove whether or not this is an advantage, but it seems reasonable to assume that a possible lengthening of the tilt motion should not be very pronounced.

In Figure 3, both maximum roll velocity and maximum lateral jerk are higher than in Figure 1 where the motions are in phase. This situation may occur with a fixed compensation ratio, which is too high for curves with very high lateral acceleration in the track plane. Reducing the tilt compensation ratio would be advantageous in this case, so that maximum tilt angle is reached at the end of the transition and not earlier. However, a general reduction may add discomfort on curves with less lateral acceleration in the track plane.

Transition curves with S-shaped curvature functions and S-shaped superelevation ramps have been investigated and vehicle response on these kind of alignments have not proven to be better than on traditional transition curves with a linear change of curvature and cant [12–14].

The present paper focuses on motion patterns according to Figure 1, but the tilt strategy is applicable also to cases according to Figure 2. An advanced tilt strategy should preclude cases according to Figure 3.



3 Comfort criteria

Comfort disturbances due to high-speed curving are of two major categories:

1. Instantaneous discomfort when horizontal curves are entered.
2. Motion sickness, which is believed to be an accumulated effect of curving.

3.1 Passenger comfort on curve transitions

From extensive tests with tilting and non-tilting trains in United Kingdom [5], P_{CT} functions for (instantaneous) passenger comfort on curve transitions have been derived as:

for standing passengers

$$P_{CT} = \max(28.54 \cdot \dot{y} + 20.69 \cdot \ddot{y} - 11.1, 0) + 0.185 \cdot (\dot{\theta})^{2.283} \quad (3)$$

and for seated passengers

$$P_{CT} = \max(8.97 \cdot \dot{y} + 9.68 \cdot \ddot{y} - 5.9, 0) + 0.120 \cdot (\dot{\theta})^{1.626} \quad (4)$$

where $\dot{y}$ is maximum magnitude of lateral acceleration in m/s^2 , $\ddot{y}$ is maximum magnitude of lateral jerk in m/s^3 , and $\dot{\theta}$ is maximum magnitude of roll velocity in degrees/s. (For low-pass filtering and other evaluation procedures, see [15].)

The P_{CT} functions express the expected percentage of dissatisfied passengers when the vehicle is running on transition curves where the lateral acceleration increases. They have been included in a European prestandard [15] of comfort evaluation.

3.2 Motion sickness

Even though motion sickness is a rare phenomenon in railway traffic, it has been reported from France [9], Great Britain [3, 5, 11], Italy [3], Japan [3, 4, 6], Sweden [7, 10] and Switzerland [8] that tilting trains sometimes cause motion sickness. Motion sickness is probably not correlated to a single curve but is rather an accumulated effect from several curves. According to ORE [3], the Japanese railways believe the problems to be associated with short transition curves. In Japan, limits of 5 degrees/s for roll velocity and 15 degrees/s² for roll acceleration have been introduced, and the problems with motion sickness have decreased, but are not totally eliminated [6].

Swedish research in this area uses net dose models where motion doses (such as integrated squared roll velocity) are accumulated over time, and where a leakage component can quantify the recovery from nausea when the vehicle is running over more straight sections of the track [16, 17]. Roll motions are believed to be a major contributor to motion sickness [10], even though vertical, lateral, yaw and pitch motions may contribute as well.



4 Optimisation algorithms

In the P_{CT} functions, eqns (3) and (4), the first term (the lateral term) is zero for low values of lateral acceleration and lateral jerk, while the second term (the roll term) is zero only for a constant roll angle ($\dot{\mathcal{G}} = 0$).

The analysis below assumes that comfort disturbances are minimised either for standing or seated passengers.

4.1 Minimising P_{CT}

For an entry transition, between a straight and a circular curve, the following cases may occur. In all cases, the train needs to “know” in advance the radius, cant and transition length (as well as the train speed).

Case Ia: There is no cant on the curve, and the lateral term is zero also for $\mathcal{G} = 0$. Lowest P_{CT} is achieved if the tilt motion is switched off (or just compensates for outward sway due to primary and secondary suspensions). This is believed to be advantageous also with respect to motion sickness since it eliminates contributions to the doses of roll motions.

Case Ib: There is cant on the curve, but the lateral term is zero also for $\mathcal{G} = 0$. Lowest P_{CT} is achieved if the tilt motion eliminates the cant angle. As in Case Ia, this is an advantage also for motion sickness.

Case II: A roll angle is required to reduce the lateral term to zero. The smallest possible roll angle is applied to reduce the lateral term to zero and to minimise contributions to doses of roll motions. The maximum tilt angle which the train can provide is not exceeded. A predefined limit for roll velocity (such as 5 degrees per second) is not exceeded. By combining eqns (1), (3) and (4), the optimal roll angle for standing and seated passengers can be calculated (see eqns (5) and (6) respectively).

$$\left(\frac{V^2}{R} \cdot \cos(\mathcal{G}) - \left(g + \frac{V^2}{R_v} \right) \cdot \sin(\mathcal{G}) \right) \cdot \left(28.54 + \frac{20.69 \cdot V}{L_t} \right) = 11.1 \quad (5)$$

$$\left(\frac{V^2}{R} \cdot \cos(\mathcal{G}) - \left(g + \frac{V^2}{R_v} \right) \cdot \sin(\mathcal{G}) \right) \cdot \left(8.97 + \frac{9.68 \cdot V}{L_t} \right) = 5.9 \quad (6)$$

where L_t is the length of the transition curve in metres.

When the optimal roll angle has been calculated, the cant angle should be subtracted and the rate of change of tilt angle can be calculated using the values of the train speed and length of the transition curve.

Case III: If the tilt angle derived in Case II exceeds the maximum tilt angle for the actual type of train, the calculation of the rate of change of tilt angle is based on this maximum tilt angle. This eliminates the risk of the saturation of the



tilt system according to Figure 3, and reduces maximum values for both lateral jerk and roll velocity.

Case IV: If the tilt angle derived in Case II or Case III generates a roll velocity (on the entry transition or the run-out transition) greater than the predefined limit, the angle is reduced to meet this requirement. This will increase the resulting P_{CT} values (increase the instantaneous discomfort), but will reduce the tendencies to motion sickness.

When these rules are applied for calculating the amount of tilt, the tilt compensation ratio will vary from curve to curve. P_{CT} will be minimised, maintaining a boundary condition for roll velocities, based on experiences of tendencies to motion sickness reported from Japan.

4.2 Reducing the doses of the roll motions further

If the predicted accumulated dose of roll motions is still too high, when the strategy in Section 4.1 is applied, there are two strategies to reduce the dose.

Strategy A reduces the limit for the roll velocity below the value used to define Case IV curves. This may reduce the amount of tilt on curves with high P_{CT} values, which then would cause even higher instantaneous discomfort.

Strategy B reduces the amount of roll motions on curves where P_{CT} values are zero or very low. Strategy B primarily affects the Case Ia, Ib and II curves. Instead of the target $P_{CT} \approx 0$, a new target is defined (such as $P_{CT} = 3\%$). The target may be adjusted until a sufficiently small dose of roll motions is predicted.

Eqns (7) and (8), for standing and seated passengers respectively, defines the roll angle $\mathcal{G}$ for Case II curves with target $P_{CT} = T\%$.

$$\left(\frac{V^2}{R} \cdot \cos(\mathcal{G}) - \left(g + \frac{V^2}{R_v} \right) \cdot \sin(\mathcal{G}) \right) \cdot \left(28.54 + \frac{20.69 \cdot V}{L_t} \right) - 11.1 + 0.185 \cdot \left(\frac{V \cdot \mathcal{G}}{L_t} \right)^{2.283} = T \quad (7)$$

$$\left(\frac{V^2}{R} \cdot \cos(\mathcal{G}) - \left(g + \frac{V^2}{R_v} \right) \cdot \sin(\mathcal{G}) \right) \cdot \left(8.97 + \frac{9.68 \cdot V}{L_t} \right) - 5.9 + 0.120 \cdot \left(\frac{V \cdot \mathcal{G}}{L_t} \right)^{1.626} = T \quad (8)$$

When these rules are applied for reducing the doses of roll, the tilt will be reduced on the curves with the best comfort and maintained at higher values for curves where the lateral term in the P_{CT} functions generates most discomfort.



4.3 Numerical examples

These numerical examples assume that the vehicle speed is 180 km/h ($V=50$ m/s) and that the tilt system saturates at a net tilt angle of 6.5 degrees. This net angle includes the outward sway of the vehicle body due to the primary and secondary suspensions.

The horizontal curves are defined in Table 1. The curves are not claimed to be representative. They are chosen just to illustrate the effect of the different tilt strategies. When doses of roll motions (degrees²/second) are calculated, the entry and exit transitions for each curve are assumed to be of the same length, and no leakage is applied (the five curves may be assumed to be close to each other).

Table 1: Track data in the numerical examples.

Curve number	Radius (m)	Cant (mm)	Transition curve (m)
1	2540	90	50
2	1500	20	80
3	1150	100	120
4	850	150	200
5	850	150	118

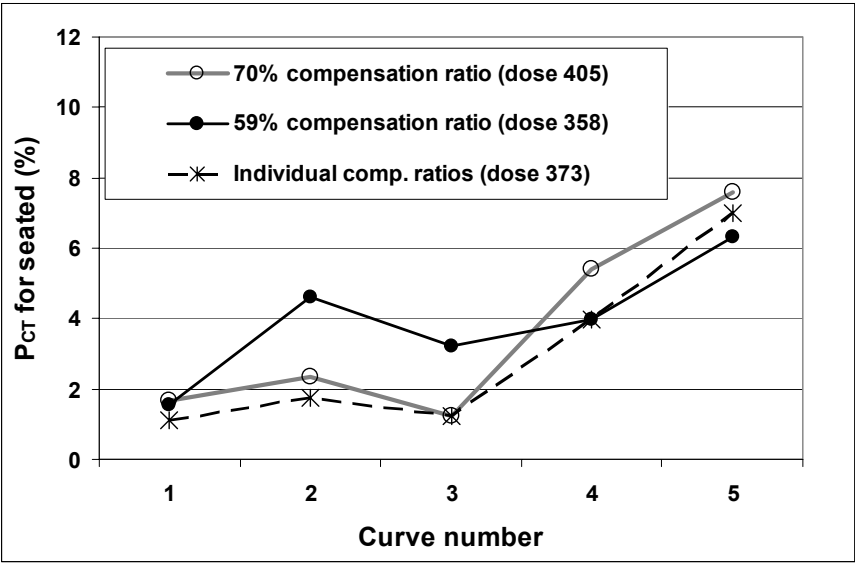


Figure 4: P_{CT} and doses of roll motions in the numerical examples.

A constant compensation ratio of 70% gives a saturation of the tilt on the transitions to curves 4 and 5. A lower ratio of 59% reduces P_{CT} on these two curves, but increases P_{CT} on curves 2 and 3, see Figure 4. With an individual

choice of compensation ratios, the P_{CT} values can be optimised for each curve, taking into account a limit of 5 degrees per second for the roll velocity (which is binding for curve number 5). If the resulting dose of roll velocity of $373 \text{ deg}^2/\text{s}$ is considered too high, a reduction of roll motions according to Strategy A will increase P_{CT} on curve 5, while Strategy B will increase P_{CT} on curves 1–3.

5 Data and computer aspects

The tilt strategies described in this paper require onboard data of the track geometry (such as curve radii, cant, lengths of transition curves and longitudinal distance between all tangent points). The data can either be the acquired from the track engineer's register of track geometry, or by motion measurements by the tilting train itself. If the train measures lateral acceleration, roll velocity and train speed, then the necessary track geometry data can be calculated and stored on the onboard computer for use on future runs.

When the train has the route data, it needs to know its position along the track. The position can be determined from a measurement of longitudinal distance, which can be calibrated by known positions of signalling transponders, GPS-measurements and/or autocorrelation with the route data.

6 Discussion and conclusions

The tilt system of a tilting train reduces the perceived lateral acceleration and lateral jerk, but increases the roll motions (among others). Therefore the tilt motion generated by computer-controlled actuators should be optimised with care, taking local track geometry and actual train speed into account.

The tilt control system that is discussed here has the following advantages. The optimal tilt angle can be calculated and arranged, taking local track characteristics into account. If it is found advantageous, the tilt motion may be started slightly before the transition curve. Unintentional tilt motions, caused by track irregularities, can be eliminated.

Furthermore, delays in the tilt motion for the first vehicle can be eliminated and the maximum tilt angle can be taken into consideration at the beginning of the transition curve (eliminating the of risk saturation of the tilt system).

In case the doses of roll motions are predicted to be too high, the amount of tilt can be reduced where instantaneous comfort on curve entries anyway is good. High values for tilt motions are maintained where the lateral motions may cause the most discomfort.

Acknowledgement

The authors thank Bombardier for permission to publish this paper.



References

- [1] Deischl, WVV. Linienverbesserungen oder gesteuerte Achsen? *Verkehrstechnische Woche*, **31(9)**, pp. 97-108, 1938.
- [2] Drechsel, A. Die Lösung der Schnellverkehrsfrage durch den kurvneigenden Kreiselwagen. *Zeitung des Vereins Mitteleuropäischer Eisenbahnverwaltungen*, **78(20)-78(21)**, pp. 377-385, 397-403, 1938.
- [3] ORE. *Body tilt coaches - Operational experience with body tilt coaches (BT)*, ORE S1037 Rp1, ORE: Utrecht, 1985.
- [4] Koyanagi, S. Ride quality evaluation of a pendulum car. *Quarterly Report of RTRI*, **26(3)**, pp. 89-92, 1985.
- [5] Harborough, P.R. *Passenger comfort during high speed curving - analysis and conclusions*, BRR TR DOS 017, British Rail Research: Derby, 1986.
- [6] Ohno, H. What aspect is needed for a better understanding of tilt sickness? *Quarterly Report of RTRI*, **37(1)**, pp. 9-13, 1996.
- [7] Förstberg, J. & Ledin, T. *Discomfort caused by low-frequency motions - A literature survey of hypotheses and possible causes of motion sickness*. VTI Meddelande 802A, VTI: Linköping, 1996.
- [8] Hughes, M. Tilt nausea is bad business. *Railway Gazette International*, **153(4)**, p. 249, 1997.
- [9] Cleon, et al. Research on motion sickness. *Proc. of World Congress on Railway Research 1999*, Tokyo.
- [10] Förstberg, J. *Ride comfort and motion sickness in tilting trains - Human responses to motion environments in train and simulator experiments*. PhD thesis, TRITA-FKT Report 2000:28, KTH: Stockholm.
- [11] Brume, M. Why are Pendolinos so nauseating? *Rail Professional*, **93**, p. 5, 2004.
- [12] Kufver, B. Realigning railways in track renewals - Linear versus S-shaped superelevation ramps. *Proc. of Railway Engineering -99*, ECS Publications: Edinburgh. ISBN 0-947644-39-3.
- [13] Kufver, B. Dynamic vehicle reactions on the Ruch type of S-shaped superelevation ramps. *Proc. of Comprail 2000*, CMP/WIT Press: Southampton, pp. 663-672. ISBN 1- 85312-598-9.
- [14] Kufver, B. *Optimisation of horizontal alignments for railways - Procedures involving evaluation of dynamic vehicle response*. PhD thesis, TRITA-FKT Report 2000:47, KTH: Stockholm.
- [15] CEN. *Railway applications - Ride comfort for passengers - Measurement and control*. ENV 12299:1999, CEN: Brussels, 1999.
- [16] Kufver, B. & Förstberg, J. A net dose model for development of nausea. *Proc. of 34th meeting of the UK Group on Human Response to Vibration*, Dunton, Essex, England, 1999.
- [17] Kufver, B & Förstberg, J. Research on the specific aspects of tilting. *Proc. of World Congress on Railway Research 2001*, Köln.



Railway car dynamic response to track transition curve and single standard turnout

J. Drożdziel & B. Sowiński

Warsaw University of Technology, Poland

Abstract

In this paper vehicle dynamic responses to clothoidal, sinusoidal and parabolic types of transition curves are considered. The responses such as vehicle motion and wheel-rail forces are significantly dependent on the type of curve. Also the dynamic interaction between the railway car and a track on a switch and a crossing of a standard turnout is considered. It was made for car-running direction “on” and “from” a switch facing point, and for nominal and geometrically imperfect track. The results imply the conclusion that specific transition curve and turnout parameters create a vital cause of dynamic interaction growth between a vehicle and a track.

Keywords: vehicle, track, dynamics, transition curve, standard turnout.

1 Introduction

In this paper, the railway car-track system dynamic interaction influenced by track transient excitations is presented. A few transition curves and a single standard turnout are taken into account. Simulation results are obtained and discussed for elastically moveable rails.

Modern railway tracks designed for high speed traffic should meet a number of requirements regarding travel comfort and safety. Designed transition track geometry changes may cause additional dynamic forces which are dependent, among others, on the shape and magnitude of these changes. During track operation the changes are additionally affected by the state of track maintenance. They are also the cause of increased local degradation of track components. For analyzing dynamic processes of this type, an appropriate simulation model was elaborated.



We have examined some aspects of the effect of track parameters variations on the system reactions. Examinations connected with the estimation of travelling comfort and safety on transition curve and a turnout constituted the basis of these investigations. In the course of applied complex studies, computer simulations were performed to assess the influence of wheelset-track system parameter changes on its dynamic behaviour. Some results of the above mentioned aspects are contained in this approach. The approach extensively examines the idea of the track parameter variations. It refers not only to disperse real values of specific parameters/dimensions regarding their nominal values but also to their deviations (e.g. nominal track gauge in relation to its real gauge). The non-symmetry phenomena which can occur in vehicle-track mechanical system are also taken into account (e.g. non-symmetrical rail profiles in the turnout zone).

2 Rail car-track dynamic system

The model description is aimed for analysing a rail car running through a transition curve and a standard turnout. The model of vehicle motion along a track consists of two interacting subsystems, namely railway car and track. Track model is subordinated to the area of vehicle-track system analysis. Railway car model (as MBS) consists of inertial solids which are connected by means of massless primary and secondary suspensions. To be exact, it consists of four wheelsets, two bogie frames (including other parts rigidly attached to the wheelset or the frame), and a body. Spring and damping characteristics of the suspensions may be introduced to the system as linear or non-linear. The test includes studying vehicle motion in relation to a reference system that moves along the track axle with the speed of undisturbed vehicle motion. Depending on the geometry of track layout, the reference system may be inertial or non-inertial.

Track model is subordinated to the analysis of dynamics pertaining to lower frequencies. The term of „lower frequencies” has been adopted from Grassi and Knothe [1] and refers to the range below 20 or 30 Hz. In the low-frequency model, the track is usually a rigid system, with the nominal layout geometry in the form of straight sections, transition curves, circular arcs, crossings and turnouts. Deviations from the nominal track dimensions are described as geometric irregularities (imperfections). Nominal track variable geometry and its deviations are kinematic excitations that influence the wheelsets of the vehicle running along the track. The deviations may also be considered while calculating geometric contact parameters and wheel-rail forces. If the lateral wheel-rail forces are significant, such as on transition curves or turnout, then elastically moveable rails may be introduced. In that case, massless or mass equivalent rail models are adopted. The models co-operate with the car wheelsets. As a result, the equivalent mass of rails under each wheel has 3 degrees of freedom (vertical, lateral, and rotation), Drożdż [2].

Assuming that the vehicle moves on a rigid track, the complete model has 27 degrees of freedom (DOF), whereas if the vehicle runs on elastically moveable rails the model has 51 DOF. The application of a more complex model ought to



be conditioned by a preliminary assessment of the significance of track elasticity, because in many cases unmoveable rigid track model is sufficient.

3 Transition curves

Let us present three types of transition curves often applied in railway track design. There are clothoidal, parabolic and sinusoidal types. As we know, transition curves are used between straight and curved track or between two adjacent curves to allow gradual change in lateral acceleration. Geometrically they are three dimensional curves with variable curvature and they also introduce variable track cant and, if necessary, gauge widening. The centre line of the transition curve has to fulfil boundary conditions at the connection points of the adjacent track component, Esveld [3]. In this approach we confined numerical results to transition between straight and curved track. Linear variation of the track curvature radius is coupled by differential-integral equation, Sowiński [4]:

$$\frac{(1 + (y'(x))^2)^{\frac{3}{2}}}{y''(x)} = \frac{\int_0^x \sqrt{1 + (y'(u))^2} du}{l} \rho \quad (1)$$

where

- x - co-ordinate measured along the track centre line $0 < x < l$,
- $y(x)$ - transition curve in plane (in horizontal projection),
- l - transition curve length,
- ρ - curvature radius of circular arc.

The result of the solution eqn (1) is a curve called clothoid. The curve gives us linear increase of centrifugal force acting on a vehicle. Also a track cant on a transition curve is introduced as linear function of variable x : $h(x) = h_0 x/l$, (h_0 – track cant on regular arc). But, the above form of cant brings undifferentiability of $h(x)$ on the curve ends. To avoid it, the curve is described using continuity conditions for $h(x)$ and demanding tangent to be zero on both ends of the curve. These conditions can be treated as boundary conditions of finite Bernoulli-Euler beam, Sowiński [4]:

$$\frac{d^4 h}{dx^4} = 0, \quad h(0) = 0, \quad h'(0) = 0, \quad h(l) = h_0, \quad h'(l) = 0$$

Finally, polynomial function of the 3rd order is obtained:

$$h(x) = \frac{h_0}{l^2} x^2 (3 - 2 \frac{x}{l}) \quad (2)$$

Now we compare the curve expressed by eqn (2) to the other curves used in track design, Sysak [5]. The first, transition curve of sinusoidal type:

$$y(x) = \frac{1}{2\rho} \left[\frac{x^2}{2} - \frac{l^2}{\pi^2} \left(1 - \cos \frac{\pi x}{l} \right) \right]; \quad h(x) = \frac{h_0}{2} \left(1 - \cos \frac{\pi x}{l} \right) \quad (3)$$



and the second, a curve of parabolic type of the fourth degree:

$$y(x) = \frac{1}{2} \frac{x^4}{6l^2\rho}; \quad h(x) = \begin{cases} \frac{2h_0}{l^2}x^2 & 0 \leq x \leq \frac{l}{2} \\ h_0 - \frac{2h_0}{l^2}(l-x)^2 & \frac{l}{2} \leq x \leq l \end{cases} \quad (4)$$

It is easy to notice, that transition curves' descriptions in the forms of eqns (2), (3) and (4) formally carry out all the conditions expressed above. Let us consider geometric differences between them. In the beginning, the changes of track curvature and track cant will be compared. Numerical calculations were performed for transition curves which are used between straight and curved track (fig. 1 and 2). The radius of track regular arc was assumed to be $R=500\text{m}$ and track cant $h_0=0.15\text{m}$.

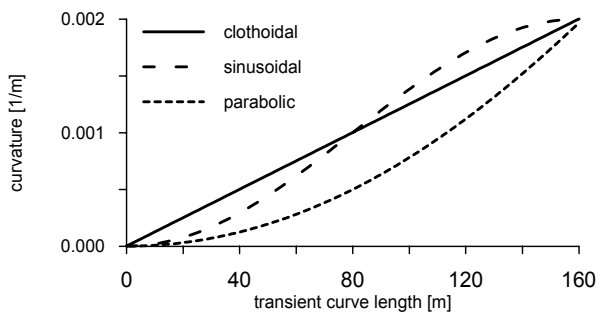


Figure 1: Comparison of track curvature for three transition curves.

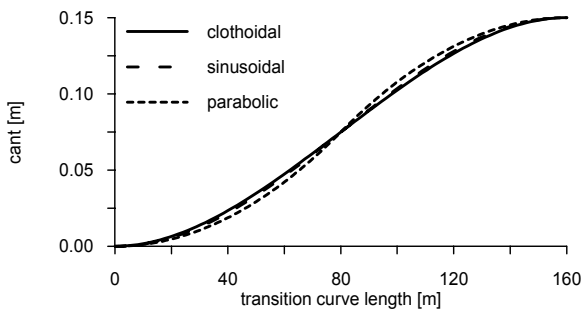


Figure 2: Comparison of cant for three transition curves.

Significant differences are visible first of all for track curvature changes. As a result of it, the differences between vehicle centripetal acceleration for sinusoidal and parabolic curves in relation to clothoidal curve amount to 0.25m/s^2 at speed of 80km/h (fig. 3). On the basis of rail passenger car simulation [6], these differences really influence the wheelsets dynamic behaviour.



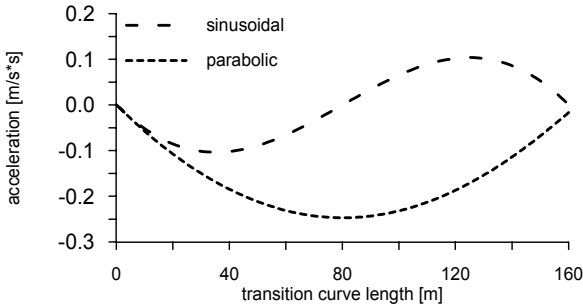


Figure 3: Differences between vehicle centripetal acceleration for sinusoidal and parabolic curves in relation to clothoidal curve.

4 Single standard turnout

For transition curves and turnouts, differences between “left” and “right” (outer and inner) rail parameters are observed. One can say that non-symmetry of geometric features occurs. This part deals first of all with modelling them. Vehicle model is to be analysed from the point of view of the influence of that track sections on the forces of interaction between the wheelset and the track and rail car motion.

Standard turnouts may be divided into three typical sections: switch, linking rails, and frog, fig. 4. If the track is analysed as a rigid system, the turnout model is the mathematical description of its geometry. The actual shape and therefore the actual parameters characterising the track in this section are the outcome of its nominal geometry and random deviations that arise during operation. It must be noted, however, that the inertial and spring-damping properties of the rails of a nominal turnout have non-symmetry characteristics as well. For example, the vertical stiffness of the rail on the frog side is greater than that of the other rail.

Starting with a track nominal geometry, the following model assumptions have been accepted, Drożdziel and Sowiński [7]:

- Unilateral gauge spreading $e(x)$ occurs in the switch and frog areas.
- In the switch area the change of a rail profile is caused by the switch blade adjacent to the stock rail, while in the frog area it is caused by the actual frog point. The vertical cross-section of a track, presented in fig. 5, clearly shows that the profile of a switch blade - 2 adjacent to the stock rail - 1, changes the potential rolling surface of the rail. The combined profile (of switch blade and stock rail) changes as the vehicle's wheelset moves, because the switch blade cross-section enlarges and, at the same time, the stock rail goes away from the nominal track centre line. Finally, switch blade profile reaches the size of a full rail profile, and a track gauge is constant and of standard value. Track parameter changes in the frog area are similar, except that rail discontinuity occurs as well.

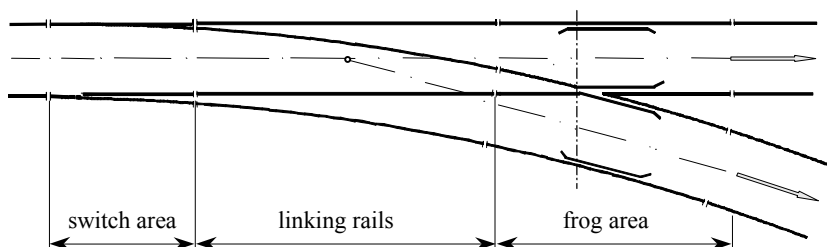


Figure 4: An overview of a single standard right-hand turnout.

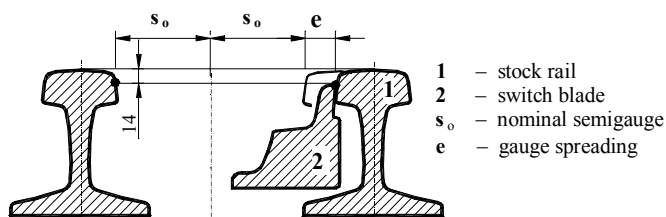


Figure 5: Unilateral gauge spreading $e(x)$ and non-symmetric rail profiles in the switch area.

- Lateral wheelset motion in the frog area is controlled by check and wing rails. The model presents them as linear-spring non-inertial barriers whose contact with wheel back flanges is the source of extra contact forces.
- When the radius of the diverging track is small, unilateral gauge spreading of a section of the inner linking rail may also be introduced. The model under analysis does not take this factor into account, because the analysis concerns diverging track radii of $R \geq 300\text{m}$.
- Apart from the ones mentioned above, there are other characteristics of the object that do not cause non-symmetry in the system and that have a significant effect on vehicle-track interactions. For example, before and after turnouts a change of lateral rail inclination takes place (at turnouts the inclination is equal to zero), and there is no cant on the diverging track.

The formation of track imperfections during operation, e.g. in the form of rail geometric deviations, is different on tracks of constant nominal parameters than in the switch and crossing area. One of the reasons for this is greater dynamic interaction between the vehicle and the track in the turnout area. The phenomenon occurs for a car-running direction “on” and “from” a switch facing point. Therefore, the description of deviations through their reduction to the track centre line equals assuming that, in the model, the deviations of both rails are identical. The above assumption is acceptable only in case of a good quality track and of constant nominal parameters. In other cases, to improve precision, independent introduction of each rail’s imperfections is suggested.

5 Simulation of rail car motion on transition curve and standard turnout

In the first part of the paper we present the comparison of leading wheelset lateral displacements and wheel-rail lateral forces for the three transition curves described above. In calculation process measured new wheels of S1002 and new rails of UIC60 profile types were used. They were approximated in pre-processing by cubic spline. All the calculations were performed for vehicle speed of 120 km/h. Dissimilarities of obtained results are evident, fig. 6 and 7.

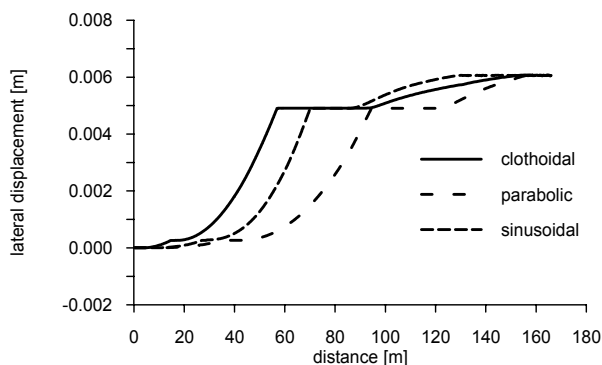


Figure 6: Lateral displacement of leading wheelset centre.

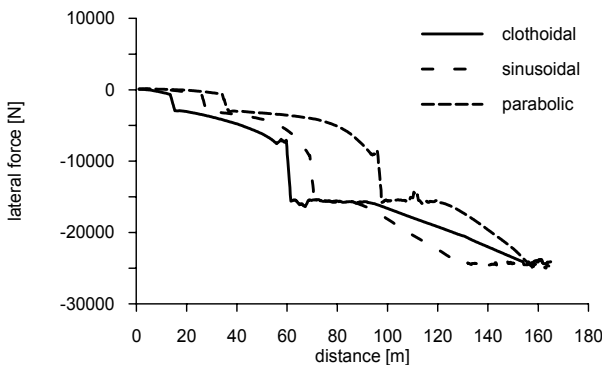


Figure 7: Wheel-rail lateral force acting on outer wheel (leading wheelset).

However, at the end of transition curves the results for displacements and forces becoming close to each other. On the graph of wheelset lateral displacements the effect of wheel profile feature is clearly marked.

Rail car passage throughout the single turnout is a typical example of vehicle-track dynamic system transient process. Numerical simulations were made on diverging track of the single standard turnout (S60-300-1:9 psd, R=300m) at speed of 40 km/h. Wheel and rail profiles are the same as for simulation runs on



transition curves. Besides, two cases were considered, that is for car running direction “on” and “from” a switch facing point. In figures 8 and 9 are shown diagrams of simulation results for running on a switch facing point direction.

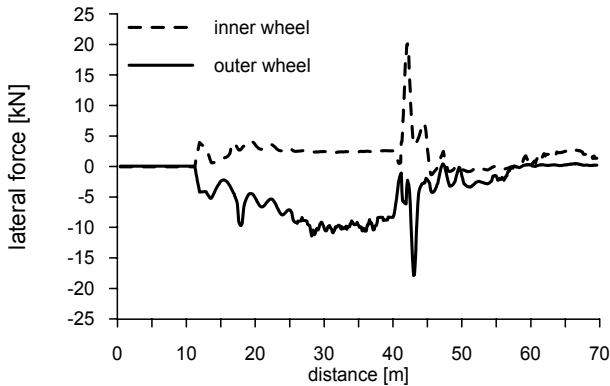


Figure 8: Wheel-rail lateral forces acting on outer and inner wheel. Running on a switch facing point.

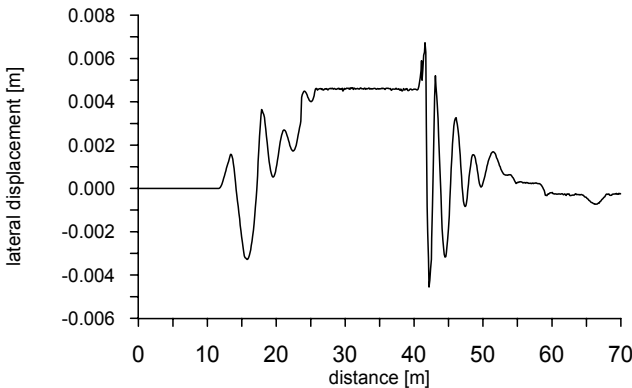


Figure 9: Wheelset centre lateral displacement.

Characteristic properties of the analysed results are the increase of wheel-rail lateral force caused by the change of rail lateral inclination prior to the switch blade and by specific track widening in the switch area. Also, notable increase of a wheelset lateral shift appears in regions of track widening (switch and frog areas). Qualitatively similar diagrams of suitable forces and displacements were obtained for the diverging track with geometric imperfections. It leads to the conclusion that the effect of turnout specific nominal geometry on vehicle dynamics is dominating. Of course, it is valid for good maintained railway track. Appropriate results for vehicle running from a switch facing point are presented in fig. 10 and 11. Additionally, to these results were supplied corresponding diagrams of 2mm narrowed track gauge in relation to its nominal dimensions. As



shown in report [6], vertical and lateral wheel-rail forces, for both runs, on and from switch facing point are not powerfully dependent on track gauge changes in tolerance limit.

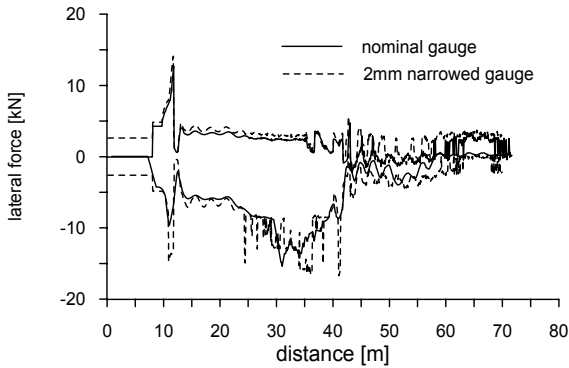


Figure 10: Wheel-rail lateral forces acting on outer and inner wheel. Running from a switch facing point.

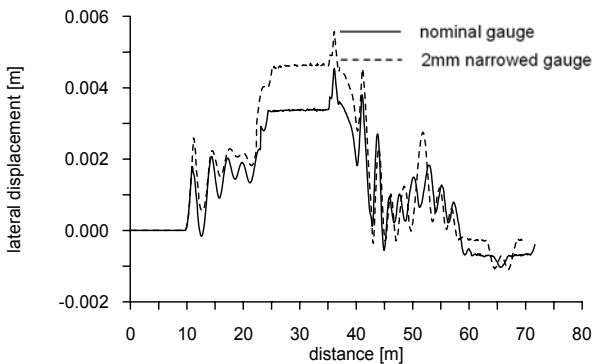


Figure 11: Wheelset centre lateral displacement. Running from a switch facing point.

Generally, they are similar in character to each other. Nevertheless, for runs from a switch direction one can observe stronger wheel-rail dynamic interaction near the switch blade, whereas for runs on a switch, stronger interaction appears in a frog area.

6 Conclusions

In the paper we have shown some possibilities of complex simulation model in application to study transient dynamic interaction in the vehicle-track system. It



has assumed that the interaction is principally affected by track geometric changes. Practical aspect of this study is the opportunity to find answer: what is the weight of that track specific properties on vehicle-track dynamic response and finally on travelling comfort and safety.

Acknowledgement

The authors wish to acknowledge the partial financial support provided by the State Committee for Scientific Research (Grant KBN No 4 T12C 006 30).

References

- [1] Grassie S.L. Knothe K., Modelling of Railway Track and Vehicle/Track Interaction at High Frequencies. *Vehicle System Dynamics*, vol 22, pp. 209-262, 1993.
- [2] Drożdźiel J., Lateral dynamics of vehicle-track mechanical system with non-symmetric distribution of parameters. *Archives of Transport*, vol.13, issue 2, Polish Academy of Sciences, Warsaw, pp. 41-56, 2001.
- [3] Esveld, K.C., *Modern Railway Track*, (2nd edition). MRT-Productions: The Netherlands, pp. 39-41, 2001.
- [4] Sowiński B., Simulation of a Railway Vehicle on Regular Track and Transition Curve. *Proc. of 14th. Proc. Intern. Conference "Current Problems in Rail Vehicles"*, Žilina, Slovakia, pp. 131-138, 1999.
- [5] Sysak J. (ed.): *Railway Roads*. Polish Publishing House (PWN): Warsaw, pp. 219-238, 1982 (in polish).
- [6] Drożdźiel J., Sowiński B. et al. *Dynamics of transient interaction between running gear of rail vehicle and track*. Warsaw University of Technology: Report of Research Project KBN, No 5 T12C 034 22, Warsaw, 2004.
- [7] Drożdźiel J., Sowiński B., Computer Simulation of Rail Vehicle-Turnout Dynamic Interaction with Mechanical Imperfections. *Proc. 7th International Conference on Computers in Railways-COMPRAIL 2000*, eds. J. Allan et al. WIT Press: Southampton, Boston, pp.673-682, 2000.



Optimization of special freight wagons with small wheel diameter

A. Rindi, D. Fioravanti, L. Pugi, M. Rinchi & J. Auciello

Università degli Studi di Firenze, Dip. Energetica "Sergio Stecco", Italy

Abstract

A better integration and interoperability between rail and road transportation is a key factor in order to reduce pollution and increase railway freight traffic.

Development of special freight wagons like "SAADKMS" for the transportation of trucks by railway is a successful solution that is meeting an increasing consensus and popularity among many European countries. In order to accelerate truck loading on wagons and reduce the limitation of normal clearance (structure/vehicle/loading gauges) it is necessary to reduce the wheel diameter as much as possible. This is not a drawback-free solution since an excessive reduction of wheel diameter involves many troubles concerning the stability of the vehicle, maximum axle load, wear of bearings and rolling surfaces of rails and axles. Also designing the braking system is very complicated because the reduced number of encumbrances available makes the placement of internal disks on the axles difficult.

In order to solve these problems a very original solution concerning wheelset, wheel profiles and more general bogie design have been applied in the development of "SAADKMS" freight wagons so the resulting vehicle is very different from the conventional one. As a matter of fact, many past experiences and know-how for conventional freight wagons are not applicable for this kind of application, so numerical simulations are very important to deeply understand the behaviour of the system and propose criteria for further optimization.

The authors of this paper have developed models on commercial multibody software in order to simulate the behaviour of the "SAADKMS" freight wagon in different conditions (stability, steering performances). Also important results such as position of the contact point or wear-number are shown to be useful parameters for further optimization of the rolling surface profiles.

Keywords: SAADKMS, wheel, rail, wear, freight, wagon, intermodal.



1 Introduction: the SAADKMS inter-modal freight wagon

SAADKMS inter-modal freight wagon has been developed by Bombardier Transportation for the transportation of trucks by railway.

Design specifications concerning encumbrances and loading procedures has lead to a drastic reduction of wheel rolling radius to about 180–190 mm.

According to Hertz contact theory [1] wheel rolling radius reduction involves a lower contact area and a higher stress of wheel material as can be shown by the following equation (simplified formulation modelling the contact between a cylinder and a planar surface):

$$p_{\max} = \frac{2F}{\pi bl}; \quad b = \sqrt{\frac{4F}{\pi l} \cdot \frac{2r(1-\nu^2)}{E}}; \quad (1)$$

where $p_{\max}$ is the maximum contact pressure, F represents the normal force between wheel and rail, l and b are respectively the length and the semi width of the contact area and r is the wheel radius.

As a consequence normal load has to be divided on a higher number of axles. Also reduced wheel rolling radius involves different behaviours concerning stability (hunting) of the bogie respect to a standard freight wagon. In fact, as stated by the simplified model of Klingel [2, 3], the hunting frequency of a single axle with a conical profile is influenced by a factor that is approximately equal to “ $r^{1/2}$ ” as shown in figure 1.

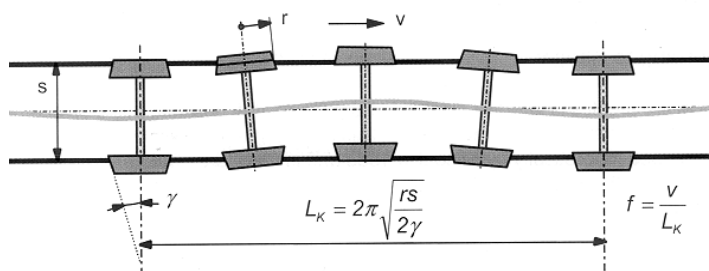


Figure 1: Axle hunting according simplified Klingel model.

SAADKMS bogie has been designed as a compromise between different engineering exigencies as visible in figure 2:

- 1) Four axles on every bogie in order to obtain a better distribution of normal loads.
- 2) Two lower frames: every lower frame is composed by two axles connected to two longitudinal elements by four spherical joints that assure an elevate angular clearance, in order to improve the steering performances of the wagon on small radius curves.



- 3) An intermediate frame is suspended on the lower frames through eight rubber deformable elements that are able to carry both vertical and longitudinal loads. Friction dampers/bumstops complete this first suspension stage.
- 4) Carbody is then suspended on the intermediate frames through a second stage of helical springs completed by friction elements that introduce damping against hunting. Longitudinal forces are transmitted by the central bearing.

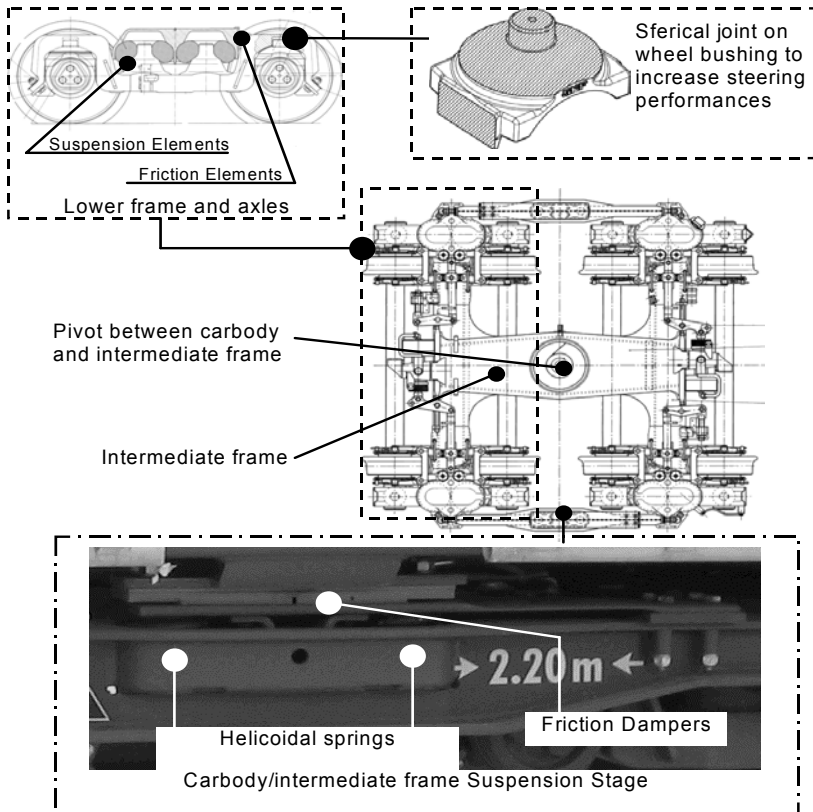


Figure 2: SAADKMS secondary suspension stage [4].

2 ADAMS model design and validation

Authors have developed a multibody model of the SAADKMS freight wagon in order to understand how different operating conditions and optimization factors may influence stability, steering and wearing of the tested vehicle. Authors have decided to develop the model using MSC.ADAMS/Rail™, one of the most



diffused and advanced tool for multibody simulation of railway vehicles. In table 1 a brief list of the most significant eigen frequencies of the multibody model of SAADKMS wagon is presented.

Table 1: The most significant eigen frequencies of the SAADKMS wagon.

MO. NB.	UND. NAT. FREQUENCY	DAMPING RATIO	REAL PART	IMAGINARY
51	2.142070E+000	2.910582E-001	-6.234670E-001	± 2.049330E+000
52	3.753565E+000	3.243317E-001	-1.217400E+000	± 3.550660E+000
53	4.498930E+000	3.471137E-001	-1.561640E+000	± 4.219200E+000
54	4.636812E+000	2.242597E-001	-1.039850E+000	± 4.518710E+000
55	5.345492E+000	2.058501E-001	-1.100370E+000	± 5.231010E+000
56	2.024580E+001	8.340102E-001	-1.688520E+001	± 1.117060E+001
57	2.024638E+001	8.326873E-001	-1.685890E+001	± 1.121130E+001
58	1.764043E+001	4.554045E-001	-8.033530E+000	± 1.570500E+001
59	1.805762E+001	4.271918E-001	-7.714070E+000	± 1.632700E+001
60	2.176496E+001	1.740655E-001	-3.788530E+000	± 2.143270E+001
61-67	2.178060E+001	1.756940E-001	-3.826720E+000	± 2.144180E+001
68-75	2.870752E+001	4.527978E-001	-1.299870E+001	± 2.559600E+001

Simulation involves simplifications and calibrations that have to be carefully evaluated in order to avoid unacceptable errors in results. Some factors as friction between rolling surfaces or the behaviour of friction dampers/end stops often introduce heavy approximation in the response of the system especially when rapid/heavy transients are involved. So the authors have decided to calibrate and validate their model by a double comparison with available experimental data kindly supplied by SBB and simulation results of a benchmark model of the vehicle developed by Trenitalia using a different software, Vampire™ from AEA Technology. First the model results have been calibrated on experimental data using record from a steering test carried on by SBB. Main features of the test are described in table 2.

Table 2: Experimental test features.

<i>Line/Track Design</i>	<i>Freight Wagon</i>
<i>Line design:</i> straight track of 100 m, blending curvature tracks (clothoid) of about 70 m, constant curvature length of 310 m with a radius of 360 m followed by a second blending track of 70 m. and a final straight path of 50 m	<i>Speed:</i> 80 km/h
	<i>Cant deficiency:</i> 0.5 m/s ²
<i>Rail Profile:</i> UIC 60 (wearied)	<i>Wheel profiles:</i> SBB 32-3 75°(wearied)
<i>Track gauge:</i> 1439.7 mm	
<i>Super elevation:</i> 135 mm	<i>Friction factor distribution:</i> 0.2 on flange 0.4 on conical surf
<i>Rail inclination:</i> 1:40	



The tested SBB vehicle was equipped with sensors in order to measure lateral and vertical forces exchanged through the wheel-rail interface.

In figure 3(a-c) some experimental are compared (lateral forces behaviour on the first axle vehicle according motion sense)

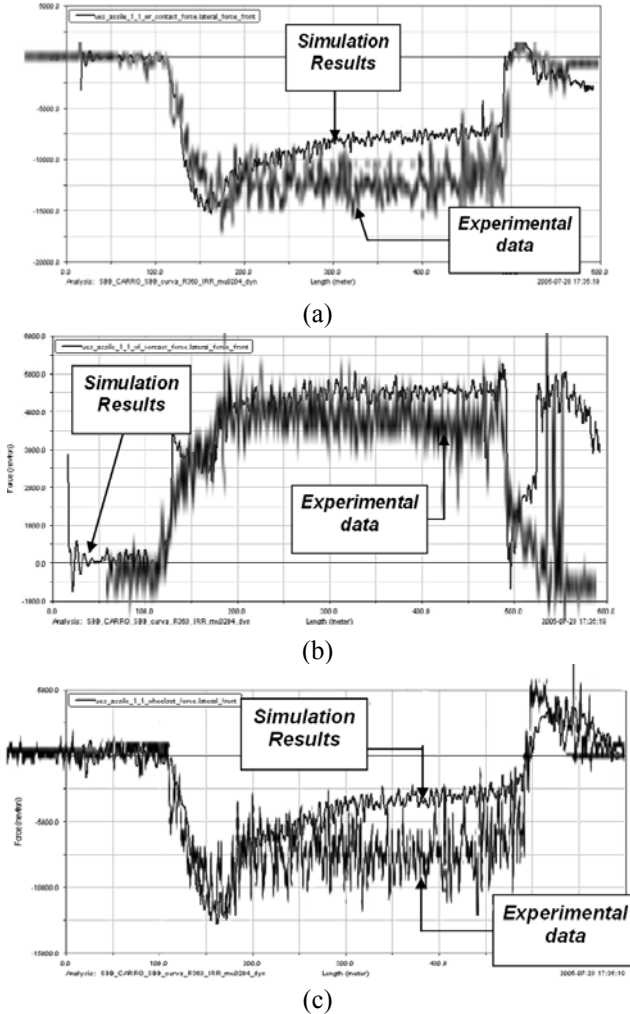


Figure 3: Lateral forces on (a) the external wheel of the first axle, (b) the internal wheel of the first axle and (c) the first axle (sum of internal and external wheel forces), comparison between experimental data and simulation results.

Simulation model show a good agreement with experimental data. Main differences are due to noise on the experimental data that were available only in paper-format and to heavy uncertainty in the modelling of both friction damper and wheel flange lubrication. In fact, it's very difficult to exactly foresee how the



lubricant on the flange influences the adhesion factor distribution along the wheel surface.

Further validation activities have involved the comparison between ADAMS model of the wagon developed by the authors and the VAMPIRE one.

Comparison between the two different models has shown a good agreement in several benchmark tests even if their internal layouts presented some differences mainly in the formulation/implementation of simulation codes. This result can be easily explained since in the comparison between simulation models, many uncertainties have been removed: for example, both models suppose the same adhesion factor distribution on wheel profiles and the same design of the track, without any difference regarding irregularities.

In table 3 we have showed the feature of one of the most significant benchmark test that the authors have carried on to compare the results of the two models: a run at various speed on a switch.

Simulations have been repeated using different wheel profiles that have been used for SAADKMS freight wagons:

- 1) The standard SBB 32-3 75° profile (the first used on SAADKMS freight wagons).
- 2) The optimized profile SBB ROLA Neu.
- 3) The optimized profile OBB SBB 32-3 FSDR3.

Table 3: Experimental test features.

<i>Line/Track Design</i>	<i>Freight Wagon</i>
<i>Line design:</i> straight track of 10 m, curve with constant curvature length of 25 m with a radius of 160 m followed by a second curve length of 25 m and a curvature radius of -160 m (opposite curvature),	<i>Speed:</i> 10/40 km/h
	<i>Cant deficiency:</i> variable according vehicle speed
<i>Rail Profile:</i> UIC 50 (NEW)	<i>Wheel profiles:</i> SBB 32-3 75° / SBB ROLA Neu / OBB SBB 32-3 FSDR3
<i>Track gauge:</i> Standard	<i>Friction factor distribution:</i> 0.2 on flange 0.4 on conical surf
<i>Super elevation:</i> 0	
<i>Rail inclination:</i> 1:20	

In figures 4(a–d) some results concerning lateral forces on the first axle during a run at 40 km/h are shown. Both the models are able to simulate the improvements in the axle behaviour with wear-optimized wheel profiles. Results also agree with experimental evidences that show the same performance improvements of optimized profiles compared with the standard SBB 32-3 75°.

Some unavoidable differences between results of the two models are mainly due to different contact formulation and friction distribution along wheel profiles.



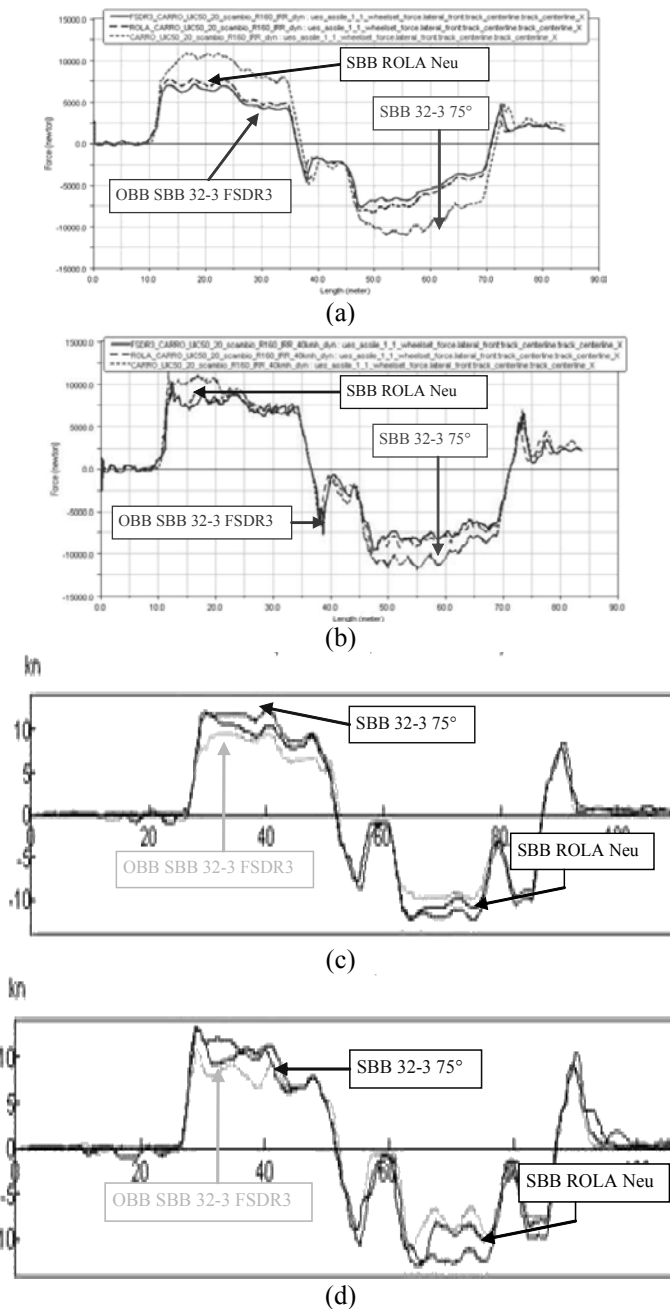


Figure 4: Lateral forces (N) on the first axle with a travelling speed of (a) 10 km/h and (b) 40 km/h (ADAMS model results). Lateral forces (kN) on the first axle with a travelling speed of (c) 10 km/h and (d) 40 km/h (VAMPIRE model results).



3 Simulation and optimization results

The most significant results concerning the optimization of the dynamical behaviour of the wagon have been first discovered during the design and the validation phase of the model and then confirmed once the model was completed by simulation results:

1. Sensitivity of ride quality and safety from friction factor of damping elements.
2. Importance of good lubrication of wheel flange and uncertainties due to an unpredictable distribution of the friction factor along the wheel profiles.
3. Wheel profile optimization.

3.1 Sensitivity from damper friction factor

In SAADKMS bogies damping is mainly introduced using friction elements. Both stability and steering performances are heavily affected from fluctuation of the damper friction factor. Experimental data from specially equipped test rig [5], visible in figure 5, show a strong dependency of the friction factor from temperature since fluctuations of few Celsius causes relative variations of friction of about 10–20%. Also a near to hyperbolic relation between wear and friction factor is clearly visible.

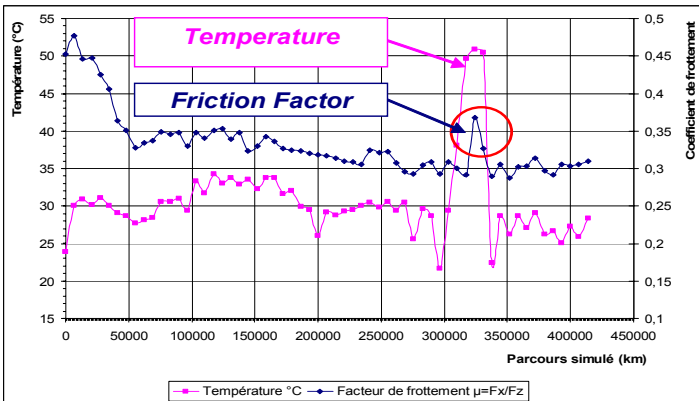


Figure 5: Friction fluctuations associated to different wear and temperature conditions [5].

During the calibration of the wagon model a medium/heuristic value of this friction factor has been chosen, however it's very difficult to predict the right value in operating conditions. So one of the hints from simulation activities it's to use more stable viscous anti-yaw damper in order to improve both stability and steering performances in a reliable manner.



3.2 Wheel flange lubrication

Actually wheel flange lubrication is adopted on SAADKMS freight wagons. Lubrication reduces flange wear and assures a better stability. In fact, as reported by simplified models often used in literature (see [1, 2, 6]), the critical ratio to avoid derailment between horizontal forces “ Y ” and vertical forces “ Q ” is negatively affected by the friction factor as visible in equation (2). In this equation “ β ” represents the geometrical flange conicity and “ φ ” the friction angle defined by (3).

$$\text{derailment occurs if : } \frac{Y}{Q} > \frac{\sin \beta - \tan \varphi \cos \beta}{\tan \varphi \sin \beta + \cos \beta} = \tan(\beta - \varphi) \quad (2)$$

$$f(\text{friction factor}) = \tan \varphi \quad (3)$$

Simulation results show a strong sensitivity in the dynamical behaviour from friction factor distribution along wheel profiles. In particular performances are affected by the friction factor in the lubricated surfaces and by the extension of these patches especially in the smoothing fillet between the conical zone of the wheel and the flange. According simulation results quality and reliability of lubrication with different operating conditions such as temperature may be critical factors that have to be deeply investigated.

3.3 Wheel profile optimization

As clearly visible in figure 7(a-d) optimized wheel profiles may increase the dynamical performances of the wagon. In particular the OBB SBB 32-3 FSD R3 has showed the best performances.

In figure 6 the standard SBB 32-75° profile is compared with the optimized one: the optimized profile has a slightly higher conicity and an increased flange clearance in order to improve steering performances of wheels.

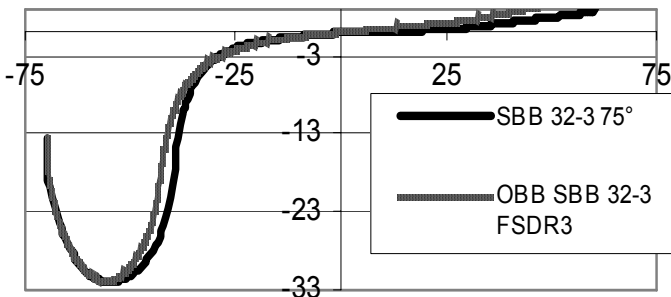


Figure 6: Comparison between standard SBB 32-3 75° profile and the optimized OBB SBB 32-3 FSD R3.



4 Conclusion and future hints

Authors have developed and validated a multibody model of the SAADKMS freight wagon and are still working in order to improve model performances and results. In particular they are still working to develop a third model of the wagon using another well known simulation tool, the Intec-Simpack™ package.

In addition, authors are applying innovative techniques based on damage/wear number criteria [7] in order to understand how further optimization for different railway profiles and inclinations may be carried on.

Acknowledgements

Research activities in the railway sector often involve the availability of technical information, experimental data, and a continuous feedback from experienced users. Cooperation is a key-word for the success of research activities. The authors wish to thanks all the technicians, the researchers and the managers that have supported in different way the activities described in this paper. In particular we remember Ing. Desideri (Trenitalia SPA), Ing. Muller (SBB).

References

- [1] J.J. Kalker, “*Three-Dimensional Elastic Bodies in Rolling Contact*”, Kluwert Academic Publishers, 1990.
- [2] Esveld, Coenraad, “*Modern Railway Track. MRT-Productions*”, Zaltbommel, The Netherlands, second edition, 2001, ISBN 90-800324-3-3.
- [3] Romano Panagin, “*La dinamica del veicolo ferroviario*” ed. Levrotto & Bella 1997
- [4] Technical documentation of Trenitalia SPA
- [5] ROLA Project Technical Documentation
- [6] A.H. Wickens, “*Fundamentals of railway vehicle dynamics*” Taylor and Francis 2003
- [7] K.J. Hewitt, “*A Model for the Prediction of Rail Wear*”. AEA Technology Rail report RR-PRJ-70012-WEAR, September 1995.



Design and simulation of railway vehicles braking operation using a scaled roller-rig

N. Bosso, A. Gugliotta & A. Somà

Politecnico di Torino, Dipartimento di Meccanica, Italy

Abstract

A 1:5 scale roller-rig has been recently realized at Politecnico di Torino for the simulation of railway dynamics. In this work it is shown that it is possible to reproduce the braking dynamics of a railway vehicle on the roller rig. For this purpose a disk-brake braking system has been designed that can be installed on the existing bogie. The inertia of the vehicle can easily be simulated adding inertia to the rollers, so that the braking distance can be evaluated on the basis of the angular displacement of the rollers. Design and optimisation of the system has been carried out with the aid of a numerical multibody model, in which different friction laws have been implemented. The simulations have been performed varying the running conditions (friction coefficient, axle load, velocity).

Keywords: wheel, rail, friction, roller-rig.

1 Introduction

A roller-rig is a test bench for railway vehicles, where the track is replaced by a pair of rollers for each wheelset of the vehicle. At Politecnico di a reduced scale roller rig (1:4 -1:5) Torino has been realized to perform dynamic simulations on a single wheelset or on an entire bogie. The realization of reduced scale models, lead to evident benefits in terms of costs of realization of the test rig and of the prototypes used to perform the tests. On the other hand it is necessary to use adequate similitude model to obtain, from the scaled results, an estimation of the behaviour of the real vehicle. Several models proposed in literature have been analysed and compared [1], and the vehicle has been designed according the similitude law proposed by Jaschinski [2].



The test rig has been initially designed to perform simulations to investigate running stability, to study the wheel-rail contact and to identify the dynamic behaviour of the vehicle. Design flexibility has soon evidenced the possible application to other field of investigation. In this work it is shown that it is possible to simulate, through a Roller-Rig, the braking dynamics for a vehicle. In this way it is possible to perform accurate analysis of variation of the friction coefficient, in a future stage of the project; the roller rig could be used to validate anti-skid systems and to test diagnostic systems which operate on railway vehicles.

In order to use the Test rig to carry on braking simulations, it is necessary first of all to realize the braking system on the bogie, furthermore it is necessary to reproduce the braked mass of the entire vehicle.

The braked mass of the vehicle can be simulated acting on the inertia moment of the rollers. Imposing that the tangential deceleration of the rollers, in braking condition, is equal to the vehicle deceleration, it is possible to calculate the inertia moment of a roller with radius R_R equivalent to a vehicle with braked mass M , as:

$$J_R = M \cdot R_R^2 \quad (1)$$

Since we want to simulate vehicles with axle load up to 25 tonnes (in 1:5 scale 200 kg/axle, it is possible to calculate an inertia equal to 6.48 kg m^2), and in order to reduce the rollers mass, the system has been designed using a gearbox with gear ratio 10, placed between the rollers and the additional inertia, as shown on figure 2 (6).

2 Numerical model

The numerical model has been realized using a Multibody code (MSC/Adams), according to the scheme shown on figure 1, where, regarding the longitudinal dynamic, a single axle has been considered.

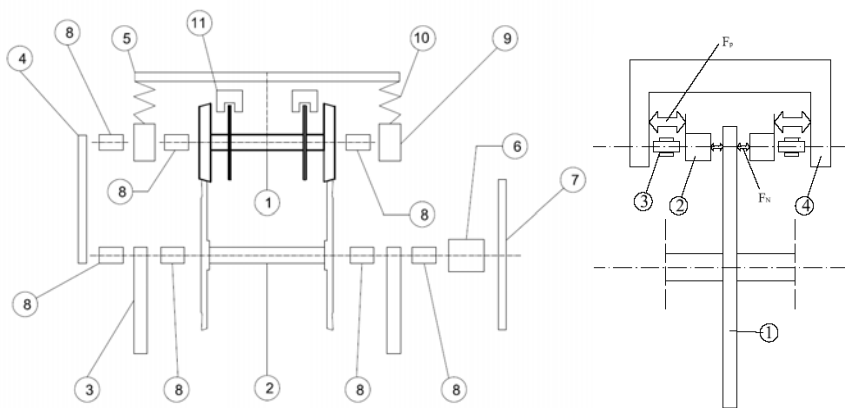


Figure 1: Multibody model scheme (left), brake system. detail.



The model is composed by the roller-rig, whose parts are the rollers supports (3) and the two rollers (2) connected by a rigid joint; the vehicle is made of one wheelset (1), connected to the bogie frame (5) by mean of the primary suspension stage (10) placed in correspondence of the axle bearing boxes (9). The bearings supporting the rollers and the wheelset have been simulated using rigid revolute joints (8). In order to reproduce the longitudinal curvature of the rollers, has been adopted the constraint scheme described in [3], composed by a mass-less dummy body (4), and by two revolute joints in order to keep fixed the distance between the rotational axis of rollers and wheelset. The braking system is made of two brake disks rigidly connected to the wheelset and by two calipers (11) connected to the bogie frame. In order to simulate the braked mass of the vehicle, this has been reported to the rollers axis, through a disk (7) and amplified by a gear-box (6). The gear box is required in order to reduce the disk mass and therefore the solicitations on the rollers supports.

The described system is brought to the test initial velocity by an electric motor connected to the left roller (where the additional inertial disk is not present) and decelerated by the braking system.

The braking system is simulated as shown on fig. 1 (right), where two pneumatic pistons (2) are allowed to slide respect the caliper (4) in lateral direction, due to the two cylindrical joints (3). The motion of the pistons is limited by the presence of the disk, by mean of a unilateral elastic force simulated as follows:

$$F_N = \begin{cases} 0 & \text{for } x \leq x_0 \\ K(x) \cdot (x - x_0) - C \cdot \dot{x} & \text{for } x > x_0 \end{cases} \quad (2)$$

where x_0 is the clearance between disk and brake pad and $x, \dot{x}$ are respectively the displacement and velocity of the piston. The braking effort is actuated by the pressure p of the pneumatic circuit, which is imposed on the system as a temporal law to simulate the braking manoeuvre.

Among each piston and the caliper as been therefore applied a pneumatic force expressed by the law:

$$F_p = p \cdot A_p \quad (3)$$

Where the frictional effects on the seals have been neglected and A_p represents the area of the piston. The friction force acting between pad and disk, which produces braking efforts, depends on the normal force expressed by eq. 3, and has been described according to a non linear and numerically stable model, as described in [4], according the equation:

$$F_f = \frac{\omega_y^s \cdot r_b \cdot \chi}{\sqrt{1 + \left(\frac{\omega_y^s \cdot r_b \cdot \chi}{|F_N| \cdot \mu} \right)^2}} \quad (4)$$

where ω_y^s is the angular velocity of the wheelset, r_b is the average braking radius (located in the center of the brake pads), μ is the pad-disk friction



coefficient (0.4) and c is a parameter (10^6) used to express the disk and pad deformation in the adhesion region of the force-velocity diagram, in order to avoid the discontinuity caused by the Coulomb's law ($\chi = \infty$).

2.1 Wheel-roller contact

The wheel-roller contact model, has been described according to [3], but only the longitudinal creep component has been considered in this work, because in case of braking it assume a dominant role in the phenomena. This force can be expressed by:

$$F_{x,DS} = - \frac{f_{11} \cdot \xi_{R,L}}{\sqrt[n]{1 + \left| \frac{f_{11} \cdot \xi_{R,L} + f_{22} \cdot \eta_{R,L} + f_{23} \cdot \phi_{R,L}}{\mu \cdot N} \right|^n}} \quad (5)$$

where the f_{ij} terms are the creepage coefficients for the linear region of the characteristics, calculated according to the Kalker linear theory [5]; ξ , η , ϕ are the kinematical creepages of the right (R) and Left (L) wheel. During braking, the most important creepage term is the one related to the longitudinal direction ξ , while the others can be neglected. A simplified (first order) expression for the longitudinal creepage read:

$$\xi_{R,L} = \pm \left(\frac{b}{\omega_s \cdot r} + \frac{\lambda}{r} \cdot \frac{r+R}{R} \cdot y \right) - \frac{\omega_s \cdot r - \omega_R \cdot R}{\omega_R \cdot R} \quad (6)$$

where b represents the gauge, r and R the contact radii of wheel and rollers respectively, λ the profiles conicity, ω_s and ω_R the angular velocity of wheelset and roller.

Equation (6) has been described for conical profiles, in case of real profiles can be extended introducing the dependence of b , r , R , λ by the lateral position of the wheelset respect the track center line (in this case defined by the pair of rollers) as shown in [3]. In this work, the analysis has been limited to the case of conical profiles, since the influence of profiles non linearity (only affect the terms among brackets of expression (6) during braking results negligible).

2.2 Wheel /rail friction

Experimental analysis performed in braking conditions show that in this case a Coulomb friction model often results erratic. For this reason several authors, develop models where the friction coefficient vary as function of the sliding velocity, normal load and creepages.

The dependence of the friction coefficient from the sliding velocity v , has been proposed at first by Kraft [6], according to results from experimental tests. The equation proposed by Kraft is:



$$\mu = \mu_0 \cdot \left(1 - 0.5 \cdot e^{-0.138/|v|} - 0.5 \cdot e^{-6.9/|v|}\right) \quad (7)$$

Where μ_0 is the static friction coefficient. The sliding velocity, in case of pure longitudinal sliding, for the wheel-roller contact (equation(6)) can be expressed by:

$$v = \xi \cdot V = \xi \cdot \omega_R \cdot R \approx (\omega_R \cdot R - \omega_S \cdot r) \quad (8)$$

The first exponential term of the equation (7), cause a sharply reduction of the friction coefficient for little values of the sliding velocity, while the second terms assure a further reduction for higher values of the sliding velocity.

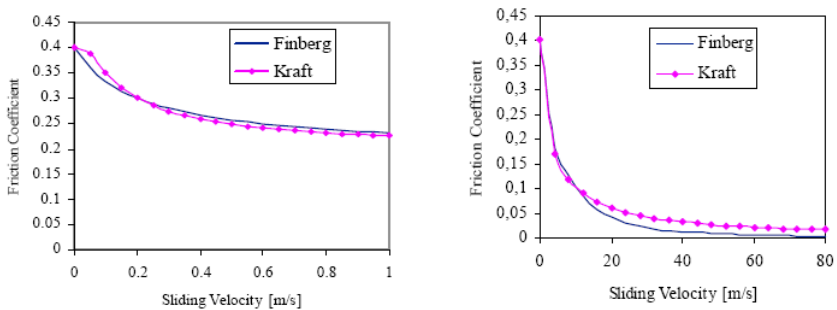


Figure 2: Behaviour of the friction coefficient according to the Kraft and Fingberg theory at low and high values of the sliding velocity.

Equation (7) was later modified by Fingberg [7], to obtain a better description of the phenomena at law sliding velocity:

$$\mu = \mu_0 \cdot \left(\frac{50}{100 + v^2} - \frac{0.1}{0.2 + |v|} \right) \quad (9)$$

The two formulations give similar results, but the Fingberg's approach is more suitable for lower sliding velocity where a more regular behaviour can be observed, Kraft's theory is more accurate at higher speed where friction coefficient evaluated using Fingberg's method vanish (see figure 2).

Periard [8] introduced two different empirical formulations as function of the sliding velocity of the wheel. The first is obtained from the experiments carried on by Poiré and Bochet of a wheel on a rail at a sliding velocity up to 20 m/s and is expressed by the relation:

$$\mu = \mu_0 \cdot \frac{1}{1 + 0.03 \cdot |v|} \quad (10)$$

The second is formulated according to the experiments of Galton [9] regarding sliding of a brake pad over a wheel during braking:



$$\mu = \mu_0 \cdot \frac{1 + 0.018 \cdot |v|}{1 + 0.097 \cdot |v|} \tag{11}$$

Both the laws show similar behaviours in the usual range of the sliding velocity (figure 3), however the friction coefficient using Galton’s method has an asymptotic trend at about 18% of the static value, while using the Periard’s method vanish.

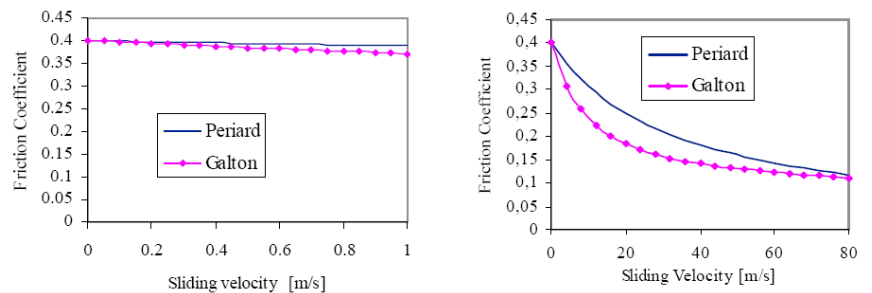


Figure 3: Friction coefficient behaviour according to the Methods of Poiré - Bochet and Galton using the description proposed by Periard.

3 Numerical analysis

Aim of this section is the analysis of the braking dynamics depending on the axle-load (12, 16 and 22.5 t/axle in real scale), of the friction coefficient (0.4, 0.2 and 0.15) and of the initial braking velocity (50, 100, 150 km/h in real scale). Numerical simulations are performed using the models described in the previous section (equation 5). All the numerical results are to be considered as referred to the reduced scale model. The results in real scale can be obtained applying the appropriate scaling factors [2]. At first a constant friction coefficient of 0.4 has been adopted, in this way has been possible to compare the results with those of analytical calculations.

Table 1: Braking times calculated analytically and numerically.

ω_0 [rpm]	ω' [rad/s]	t_1 - t_s [s]	t_s [s]	t_a [s]
(v1) 1800	2.42	0.04	1.04	1.04
(v2) 3000	23.36	0.40	1.40	1.40
(v3) 6000	75.72	1.31	2.30	2.31
(v4) 12000	180.44	3.11	4.12	4.11

The vehicle has been braked starting from different initial velocity as shown on figure 3, After 4 seconds a pressure has been applied in order to reach the maximum braking force (imposed equal to 100 N) after 1 second.

The vehicle used for the simulations has rollers inertia equal to 0.1 kg m² and an average braking radius *r_b* equal to 0.0725 m. As shown on figure 4, up to a

certain time t_1 , the brake effort develops with increasing pressure and force, after the time t_1 the normal load reach its maximum (in this case equal to 100 N).

In Table 1 are indicated the rollers velocity at time t_1 , the time interval in which the braking took place at constant force (t_1 - t_s), the total brake time evaluated using the numerical (t_s) and the analytical model (t_a).

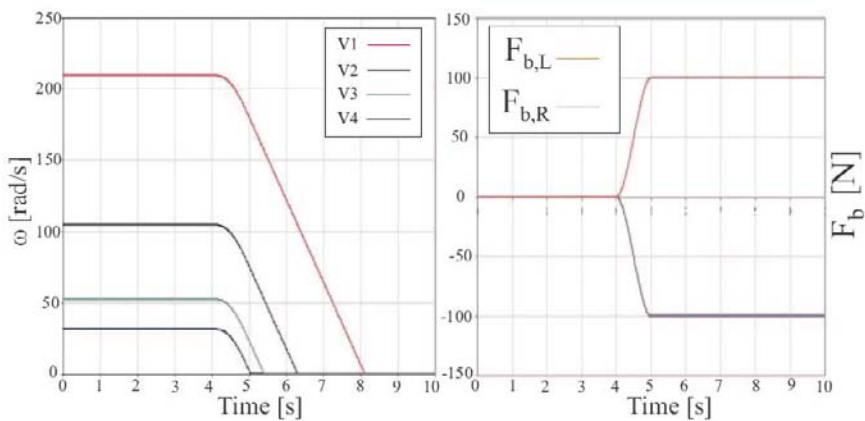


Figure 4: Rollers deceleration (left). Braking force behaviour (right).

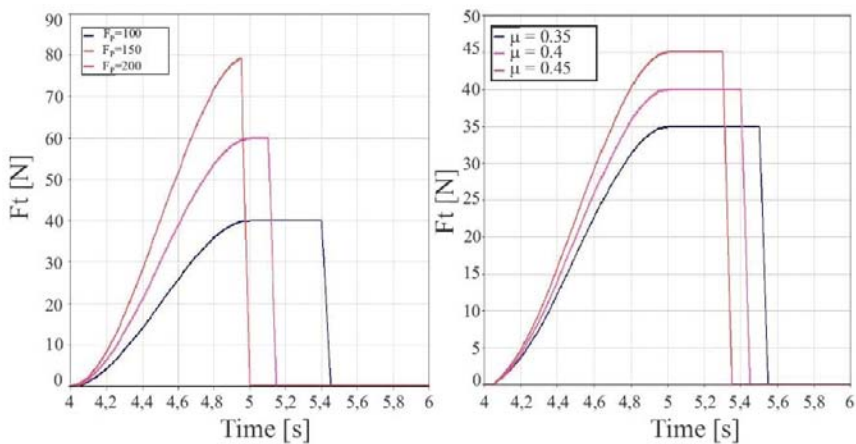


Figure 5: Friction force behaviour versus time as function of the normal load (left) and of the friction coefficient (right).

In order to evaluate the effect of the normal load and of the friction coefficient, a series of simulation has been performed starting from the velocity of 100 km/h in real scale, modifying the pressure on the calipers (100, 150, 200 N) with a friction coefficient of 0.4 as shown in figure 5 (left). Then (figure 5, right)



the normal load has been kept constant (100 N) and the friction coefficient has been modified.

Previous analyses don't consider the variation of the friction coefficient during braking. Next Fingberg's and Periard's theories will be introduced in order to evaluate this effect.

3.1 Comparison between different adhesion laws.

In this section the constant friction coefficient model will be compared with the models proposed by Periard (equation (13)) and Fingberg, according to equation (11). Fingberg's law is more suitable for low sliding velocity, while Periard's one is better at higher velocity.

Considering a braking manoeuvre with constant torque, the friction forces behaviour, shown on figure 6 (left), can be distinguished in three regions. In a first phase the friction force increase linearly together with the longitudinal creepage, during this phase the contact area is characterised by full adhesion. The second region shows a non linear increasing trend of the friction forces. This is due to the formation, in the contact patch of a portion where sliding occurs. The sliding portion gradually increases up to saturate the entire contact patch; when the entire area is in sliding condition, the maximum value of the tangential force is reached. This condition is also known as optimal adhesion condition.

Further increasing of the creepage, produce different effect depending on the considered model. Using the constant friction coefficient model, the maximum friction force remains constant.

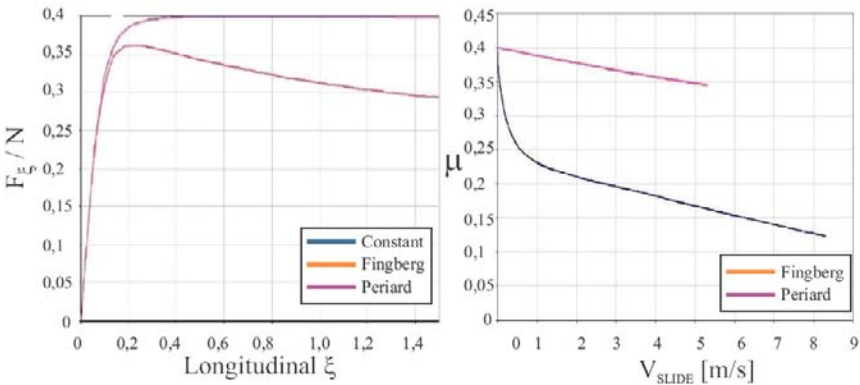


Figure 6: Comparison between different adhesion laws ($V=100$ km/h, $N=12$ t/axle): adimensional longitudinal creep force versus creepage (left); friction coefficient versus sliding velocity.

With Fingberg's and Periard's models, if the creepage increase over the maximum of adhesion, the friction force reduces. Periard's model shows a little reduction (for small creepage) respect the constant model, while Fingberg's model predict an important reduction in the same conditions.



Table 2: Comparison between different adhesion laws: N = 96 kg, V = 100 Km/h.

Law	Cb [Nm]	$F_{\varepsilon} / \mu N$ [$\sqrt{}$]	t_{stop} [s]	s_{stop} (1:5) [m]
Fingberg	32.46	0.984	4.85	43.44
Periard	35.61	0.99	4.54	41.43
μ constant	35.61	1.00	4.54	41.43

Analysing in detail the friction force behaviour, a different creepage value in correspondence of the optimal adhesion condition can be observed, equal to 0.27% using Periard’s model, and 0.19% with Fingberg’s.

The different value of the braking force lead obviously to higher stop time and distance with the Fingberg’s model, as indicated on table 2.

As can be seen on figure 6 (right), the friction coefficient in full sliding condition is 0.345 with Periard’s law and 0.134 using Fingberg’s. This is related to the large discrepancies in the braking times and distances predicted by the two methods.

Using different methods, different peak values of the sliding velocity are reached starting from the same condition (initial velocity, braking torque).

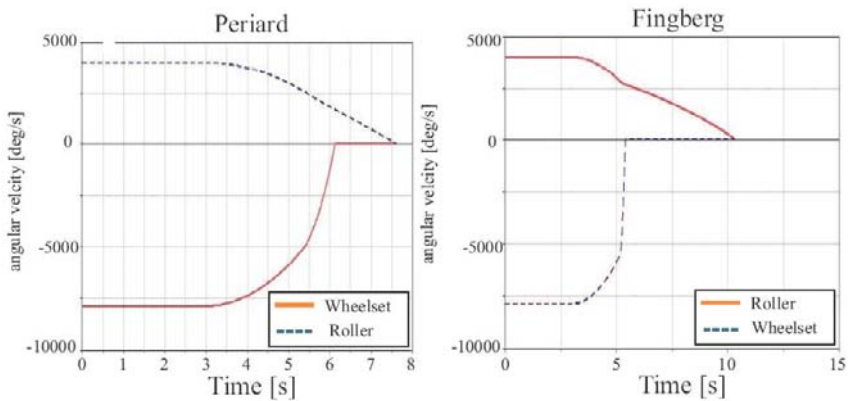


Figure 7: Wheel (solid) and roller (dashed) deceleration using Periard’s (left) and Fingberg’s (right) theory.

The time histories of the wheel and roller angular velocity, using the two approaches, shown on figure 7, allow a direct observation of the braking times.

According the model of Fingberg, the braking time results to be 7.32 s, while using Periard’s only 4.61 s are required to stop the vehicle (data in 1:5 scale). This means braking distances of 49.39 and 41.4 m using the two models.

Figure 7 show the difference of the behaviour on roller respect to what can be expected on rails: when sliding occurs, the wheelset stops while the rollers continue their rotation, which represent, on rail, the longitudinal sliding of the vehicle.

3.2 Effect of the initial velocity

The dynamical behaviour during braking is strongly influenced by the initial velocity of the vehicle, due to the dependence if the friction coefficient by the sliding velocity.

Therefore, several simulations are performed on the previous analysed model (axle load 12 t: 96 kg in 1:5 scale) modifying the initial velocity: 50,100 and 150 km/h (corresponding to angular velocity of the rollers equal to 34.5, 69 and 103.5 rad/s in 1:5 scale).

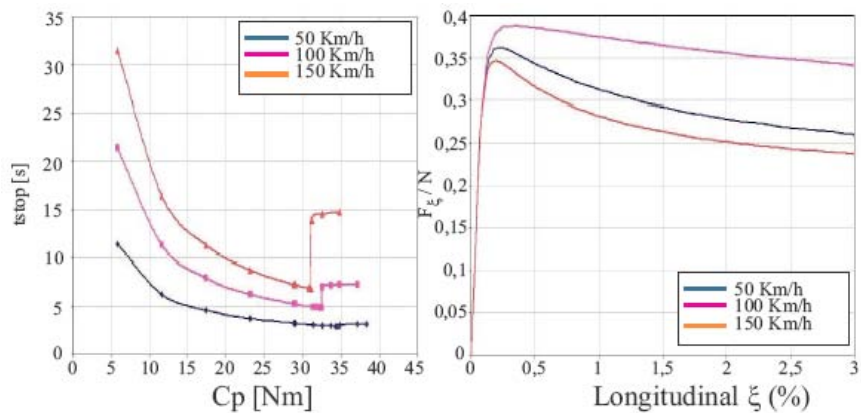


Figure 8: Stop times for different initial velocity as function of the brake torque. Adimensional creep force for different initial velocity.

Increasing the initial velocity, the maximum value of the sliding velocity increases and subsequently the friction coefficient decreases.

Therefore, performing the test with different values of the normal force acting on the brake pads (defining the brake torque), sliding occurs for high speed braking at high torque values (figure 8, left). Further increasing of the normal pressure is not useful for the stop time and distances.

On table 4, braking times and distances are indicated for different initial velocity and braking torque. The darkened fields in the table show an increment of the brake distance due to an increased torque; this means that sliding of the wheel has occurred.

Considering the longitudinal force as function of the creepage, can be observed that increasing the initial velocity, the optimal adhesion condition moves towards lower creepage values, as shown on figure 8 (right). This consideration has been quantified on table 4 and is important in order to design anti skid systems.



Table 3: Comparison between different adhesion laws: N = 96 kg.

Cb	V ₀ = 50 Km/h		V ₀ = 100 Km/h		V ₀ = 150 Km/h	
	t _{stop} [s]	s _{stop} [m]	t _{stop} [s]	s _{stop} [m]	t _{stop} [s]	s _{stop} [m]
[Nm]						
5.8	11.38	42.3	21.43	147.7	31.52	316.3
11.6	6.28	26.4	11.34	84.5	16.38	174.3
17.4	4.61	21.0	7.98	63.37	11.33	126.9
23.2	3.77	18.2	6.30	52.7	8.815	103.1
29	3.27	16.5	5.29	46.2	7.30	88.7
30.74	-	-	-	-	6.96	85.5
30.97	-	-	-	-	6.916	85.1
31.03	-	-	-	-	6.906	85.0
31.09	-	-	-	-	13.93	150.5
31.32	3.12	15.9	4.98	44.3	-	-
31.90	-	-	4.93	43.9	-	-
32.25	-	-	4.88	43.6	-	-
32.36	-	-	4.86	43.5	-	-
32.46	-	-	4.85	43.4	-	-
32.48	3.05	15.7	7.02	47.6	14.46	157.7
33.64	2.99	15.5	7.22	47.9	-	-
34.22	2.96	15.34	-	-	-	-
34.45	2.95	15.3	-	-	-	-
34.68	2.93	15.26	-	-	-	-
34.74	2.93	15.4	-	-	-	-
34.8	3.13	15.3	7.27	49.4	14.73	161.3

Table 4: Comparison between different initial velocity: N = 96 kg.

V ₀ [Km/h]	ξ _{max} [%]	F _{ξ,Max} [N]	F _{ξ,Max} /μN
50	0.255	181.97	0.99
100	0.197	170.03	0.984
150	0.17	162.53	0.97

3.3 Brake at low friction coefficient

In the previous section a static friction coefficient of 0.4 has been considered, this value can be conventionally used in case of dry contact. It is well known that environmental effects (such as rain, ice, various contaminations) degrade the wheel rail friction. In this section different case are compared: normal condition (μ₀=0.4), wet (μ₀=0.2), and contemned contact (μ₀=0.14).

Table 5: Optimal adhesion condition, different μ₀.

μ ₀	ξ _{max} [%]	F _{ξ,Max} [N]	F _{ξ,Max} /μN
0.4	0.197	170.03	0.984
0.2	0.105	87.23	0.98
0.14	0.08	62.67	0.99



As shown on table 5, in reduced friction conditions, the optimal adhesion moves towards lower creepage values. This means that, in order to avoid skidding, it is necessary to reduce the braking torque.

Furthermore, as can be observed on figure 9, using a model with variable friction coefficient (Fingberg), and increasing the creepage, the friction coefficient decreases with the same law for each values of the initial friction.

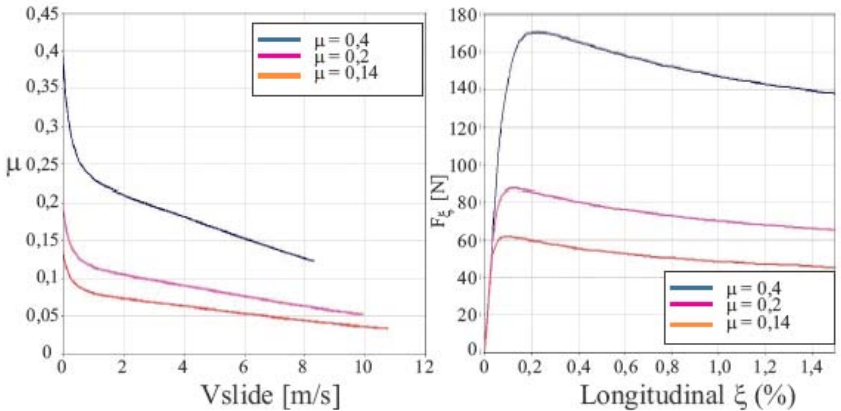


Figure 9: Friction coefficient (left) and creep force (right) for different surface condition: dry friction ($\mu_0=0.4$), wet (0.2) and contaminated (0.14).

This situation appear to be un-realistic, in fact it is difficult to believe that physical laws valid for direct contact between two elastic solids (wheel and rail) are not significantly altered by the presence of a third element among them. The problem could be further emphasized in case of long last at high sliding velocity. Therefore the analyses here illustrated are to be considered as a starting point, to be modified on the basis of further experimental tests. Test performed on roller rig could be useful for a better understanding of this phenomenon, due to the possibility to easily change and identify all the parameters.

3.4 Effect of the axle load

Finally the influence of axle load on brake dynamics is considered. Three load conditions are simulated: 12, 16 and 22.5 t/axle, corresponding to 96,128 and 180 kg in 1:5 scale.

Simulations are performed for different initial velocity values as shown in table 6, where are reported: the torque required to brake in optimal adhesion condition (C_b), deceleration, brake time and distance.

As can be noticed from figure 10 (left), the creep force is linearly proportional to the normal load except for a little variation of the optimum of adhesion (towards higher creepage values for higher loads).



Table 6: Effect of the axle load.

V_0 [m/s]	M [kg]	Cb [Nm]	t_{stop} [s]	s_{stop} [m]	$\ddot{x}$ [m/s ²]
6.21	96	34.74	2.93	15.25	3.7
	128	46.40	2.92	15.2	3.72
	180	66.58	2.93	15.23	3.71
12.42	96	32.47	4.84	43.44	3.46
	128	43.21	4.84	43.44	3.47
	180	61.54	4.87	43.64	3.44
18.63	96	31.03	6.9	84.97	3.3
	128	41.24	6.9	84.88	3.31
	180	58.6	7.0	85.6	3.28

This is more evident analyzing the behaviour of the brake torque in condition of optimum of adhesion as function of the axle load as shown in figure 10 (right).

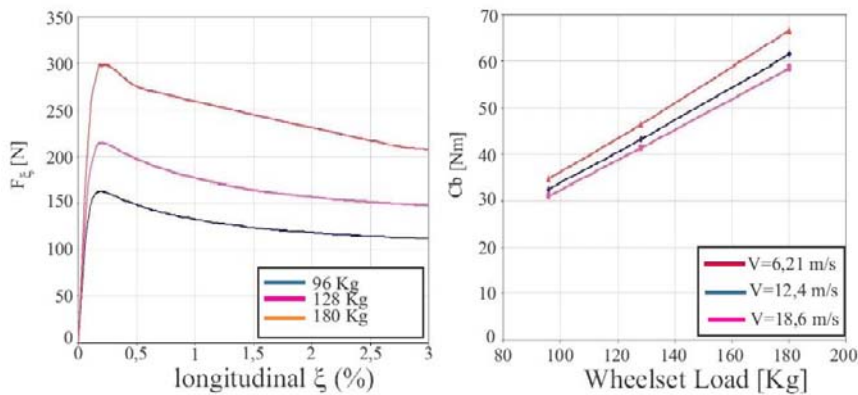


Figure 10: Braking force behaviour (Initial velocity 100 Km/h), as function of the creepage for different axle load (left), and torque in optimal adherence condition as function of the axle load for different initial velocity (right).

4 Design

The brake system is composed by two disks mounted on each wheelset (fig 11) fixed to the wheel by mean of two cylinders (2 and 3 in figure 11). The brake effort is exerted by two motorcycle type calipers (Brembo P32G) acted by air instead of oil. Experimental tests performed on this caliper demonstrate the possibility to achieve a linear regulation in the range 1-8 bar with braking efforts capable to reproduce the behaviour of the real vehicle (in scale). The calipers have been connected to the bogie frame with a beam system which integrates an estensimetric system to measure the braking force.

In order to measure the creepage, in the centre of the wheelset has been placed an incremental encoder.

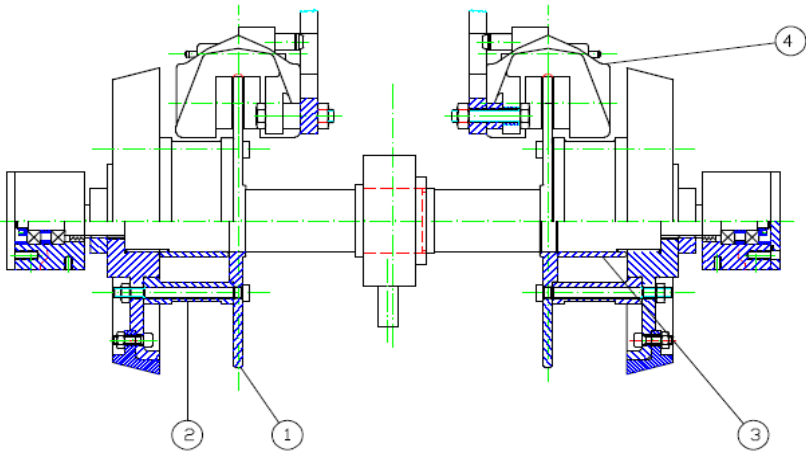


Figure 11: Drawing of the wheelset with brake disk(1) and calipers (4).

5 Conclusions

The work illustrate how is possible to use a Roller Rig in order to perform braking simulation on railway vehicles. In the first part of the work several friction laws, taken from literature are applied to a numerical model of a suspended wheelset on roller-rig in order to investigate braking dynamics.

The influence of different parameters is considered: friction coefficient, axle load, braking force and results of different theories are compared.

It is clear that different braking distance are predicted depending on the theory used, furthermore the behaviour of the friction coefficient in some conditions could not be fully explained in physical terms (figure 9).

For this reason the use of a test stand, as a Roller-Rig, under well known environmental condition can be useful to improve the modelling of the friction coefficient behaviour during braking.

The simulation has been used in the design phase of the braking system, to be applied to the existing Roller-Rig. Design has been performed in order to be able to simulate braking manoeuvre of vehicles with axle load up to 25 t/axle, and speed up to 400 km/h.

The experimental tests could be used both to investigate the friction coefficient behaviour, but also to develop more efficient anti-skid systems.

References

- [1] N. Bosso, A. Gugliotta, A. Somà, "Comparison of different scaling techniques for the dynamics of a bogie on roller rig"- Vehicle System Dynamics 37 (SUPPL.), pp. 514-530, 2003.



- [2] Jaschinski A., 1990, "On the application of similarity laws to a scaled railway bogie model.", DLR Institut für Dynamik der flugsysteme, Oberpfaffenhofen (Germania).
- [3] N. Bosso, A. Gugliotta, A. Somà, "Introduction of a wheel-rail and wheel-roller contact model for independent wheels in a Multibody code." – ASME/IEEE Joint Rail Conference , Washington, DC, 22-24 April 2001.
- [4] "Dry friction models", J.B. Ayasse, S. Iwnicki, H. Chollet, poster paper, XVIIth IAVSD Symposium, Lyngby, Denmark, 2001.
- [5] Kalker, J.J.: Three-dimensional elastic bodies in rolling contact, Kluwer Academic Publisher, 1990.
- [6] K. Kraft, "Der Einfluss der Fahrgeschwindigkeit auf den Haftwert zwischen Rad und Schiene.", Archiv für Eisenbahntechnik 22, 58-78, 1976.
- [7] U. Fingberg, "A model of wheel-rail squeal noise." Journal of sound and vibration 143, 365-377. 1990.
- [8] F. Periard. "Wheel-rail noise generation: curve squealing by trams." PhD thesis, TU Delft 1998.
- [9] G. Galton. "The action of brakes. On the effect of brakes upon railway trains." Engineering 25, 469-462. 1878.
- [10] D.J. Thompson, A.D. Monk-Steel, C.J.C. Jones, P.D. Allen, S. S. Hsu, S.D. Iwnicki, "Railway Noise: curve squeal, Roughness Growth, Friction and Wear", Rail research UK, June-2003.



This page intentionally left blank

Study on vertical dynamic vehicle-track interactions using the TRADYS test facility and computer simulation

M. Miwa¹ & A. Yoshimura²

¹*Technology Research and Development Department,*

General Technology Division, Central Japan Railway Company, Japan

²*School of Computer Science, Tokyo University of Technology, Japan*

Abstract

It is important when studying vehicle-track dynamic interactions to make clear the characteristics of many physical and/or structural parameters of a track. However, it is difficult to estimate them with accuracy only through measurements under a real train load on site. In this paper, we describe an approach using two tools of a test facility and computer simulation. The Track and Structural Dynamics Simulator (TRADYS) is a unique mechanical vibration exciter equipped with hydraulic actuators between a train and a track. The train has the same masses and suspensions as the latest Japanese high speed railway vehicle. We can control arbitrarily the vertical relative displacements between rails and wheels through the directions inputted to the actuator. These relative displacements are supposed to represent a railway track irregularity. In the experiments, dynamic wheel loads, displacements and accelerations of each part of a track are measured as the outputs. In addition to the experiments, computer simulation is carried out. A theoretical model for this special test facility including the actuator is constructed. In this paper the measured physical values and computed solutions are compared, and the accuracy of the parameter values of the track structure components is discussed.

Keywords: vehicle-track interaction, dynamic model, test facility, computer simulation, parameter estimation.



1 Introduction

We have researched not only vehicle-track dynamic interactions but also various railway technological issues using three instruments and combining them effectively as shown in Fig. 1. In the disciplines of vehicle-track dynamics, characteristics of dynamic wheel load, ground vibration and effects of unsprung-mass of a vehicle determined before with the high-speed experimental train “300X” in Miwa and Nakamura [1]. To enhance these researches aiming at more improvement of running stability and riding comfort, reduction of ground vibration and track maintenance, we introduced a unique test facility and developed a computer simulation program as described below.

The test facility in this paper generates dynamic load at the fixed point on a track while the real train works as a moving load. In the former, we can change load arbitrarily to estimate track parameters conveniently. The parameters are estimated with the computer simulation which simulates vehicle/track behavior under dynamic load at the fixed point. Then moving load is simulated with the estimated parameters and its accuracy is verified with field data.

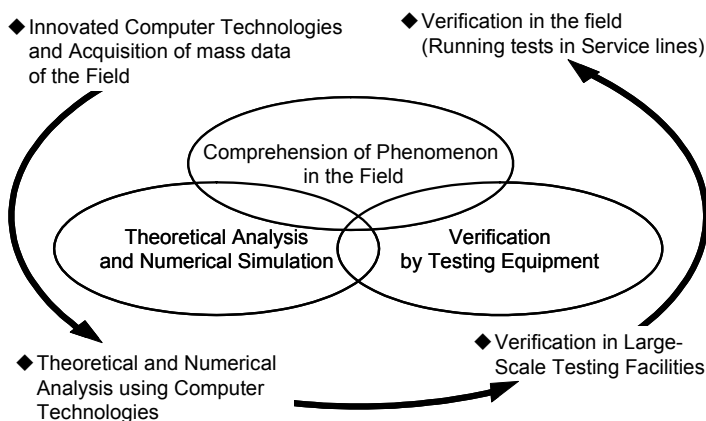


Figure 1: Basic cycle for R&D of railways.

2 Track and Structural Dynamics Simulator, TRADYS

The full view of the test facility, Track and Structural Dynamics Simulator (TRADYS), is shown in Fig. 2. Its total weight and bogies are the same as the real vehicle, the Series 700, of the Tokaido Shinkansen. The leftmost axle in Fig.2 is replaced with the vibration exciter, and it is called as the Dynamics Testing Bogie. Its detailed view is shown in Fig. 3. A pair of hydraulic actuators and dead-weight which are equivalent to the unsprung-mass of the vehicle are set here substituting for the original axle. The unsprung-mass is changeable from 1.4 to 2.3ton while its value of the Series 700 is approximately 1.8ton. Actuators are mounted on rails through jigs of the wheel tread shape, and they give relative



displacement between rails and wheels (unsprung-mass) forcibly at the fixed point. The displacement given by the actuators represents track irregularities, and therefore, the excitation force is generated as a result of vehicle-track interaction based on cause-and-effect relationship (see Fig. 4). The excitation force is considered to equivalently correspond to the one due to a moving vehicle on a track irregularity. In addition, displacement is given with regular waves, such as sine-waves, and random waves. Therefore it is easy to estimate parameters of track components by changing input waves arbitrarily.

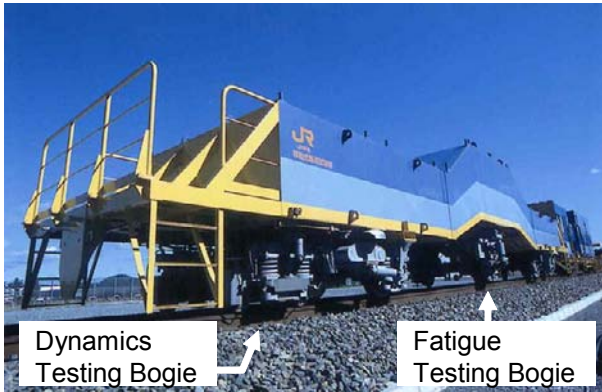


Figure 2: Track and Structural Dynamics Simulator, TRADYS test facility.

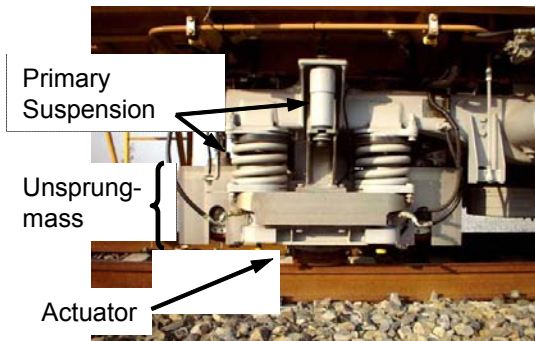


Figure 3: Dynamics testing bogie of TRADYS.

3 Mathematical model for the TRADYS test facility

Figure 4 shows the mathematical simulation model of the TRADYS test facility. The vehicle model consists of a rigid carbody, a rigid bogie and a rigid wheel. They have the same masses and suspensions as the latest Japanese high speed railway vehicle. The track consists of a single rail supported discretely by sleepers. The sleeper is modeled as a lumped mass. The pad spring and ballast spring are considered to be one-dimensional elements. The railpad works as a



viscoelastic foundation with linear elastic stiffness and viscous damping. The ballast is assumed to consist of several layers (usually three layers) having each different effective mass. They are also connected by linear springs and viscous dampers in parallel. The roadbed is fixed to a ground.

In the following, a brief description of the equations of motions of the above dynamic models of a vehicle and a track with the constraint on the motion due to the forced relative displacement is given.

The equations of motion of the vehicle model can be written as

$$\mathbf{M}_V \ddot{\mathbf{u}}_V(t) + \mathbf{C}_V \dot{\mathbf{u}}_V(t) + \mathbf{K}_V \mathbf{u}_V(t) = -P_{dyn}(t) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (1)$$

where $\mathbf{u}_V(t)$ is the displacement vector of the vehicle model (Yoshimura [2]). $\mathbf{M}_V$, $\mathbf{C}_V$, $\mathbf{K}_V$ are the mass, damping and stiffness matrices of a vehicle, respectively. $P_{dyn}(t)$ is an external dynamic force acting between a rail and a wheel due to an excitation by the actuator.

We focus on the vibration with predominant components in the frequency range from 0 Hz to 100 Hz caused by the vehicles running on tracks on the ground surface, so the rail is modeled as a simple Bernoulli-Euler beam. In the finite element analysis the rail beam is divided into N elements ($N+1$ nodes), as shown in Figure 4. The length of each element is equally h . The Bernoulli-Euler beam element is a two-node element with two degrees of freedom for each node. The nodal coordinates of the element is described as the vector form:

$$\mathbf{u}_i = (u_i, h\theta_i, u_{i+1}, h\theta_{i+1})^T \quad i = 0, 1, \dots, N-1 \quad (2)$$

where u and θ are the vertical displacement and the rotational angle respectively, i and $i+1$ are the global node numbers at each side of the i th element.

The equations of motion of the track model are obtained as follows.

$$\mathbf{M}_R \ddot{\mathbf{u}}_R(t) + \mathbf{C}_R \dot{\mathbf{u}}_R(t) + \mathbf{K}_R \mathbf{u}_R(t) = P_{dyn}(t) \begin{pmatrix} \mathbf{a} \\ 0 \end{pmatrix} \quad (3)$$

where $\mathbf{M}_R$, $\mathbf{C}_R$, $\mathbf{K}_R$ are the mass, damping and stiffness matrices of the track model respectively. $\mathbf{a}$ is a vector corresponding to x_w which represents the position of a wheel on a rail and defined as follows, when a wheel is on the i th finite element.

$$\mathbf{a} = (0, \dots, 0, N_1(\xi), N_2(\xi), N_3(\xi), N_4(\xi), 0, \dots, 0)^T$$



By differentiating the constraint eqn (5), we get

$$\dot{u}(x_w, t) + \dot{r}(t) = \dot{y}_w(t) \tag{6}$$

However, one of the difficulties with reducing the index in this way is that the numerical solution of the resulting system needed no longer satisfy the original constraints exactly. So in practice, we have adopted the following differential equation obtained by combining eqns (5) and (6).

$$2\alpha(\dot{y}_w(t) - \dot{u}(x_w, t) - \dot{r}(t)) + \beta^2(y_w(t) - u(x_w, t) - r(t)) = 0 \tag{7}$$

where α, β are adjustable constants and their optimum values are to be determined by trial and error, observing the numerical results of the simulation.

As a result, it is required for us to solve the simultaneous differential eqns (1), (3) and (7). In this study we applied the well-known Newmark- β method, one of the direct integration methods, to solve the problem. The above constants α, β were taken to $\alpha=1.0, \beta=1.0$ as optimum values by which numerical stability is attained successfully.

5 Examples of experiments in the TRADYS

The TRADYS and test track are installed in the Komaki R&D Complex (Masuda [4]) of Central Japan Railway Company. All kinds of concrete sleepers which are used in the Tokaido Shinkansen Line are laid in the test track. Test results of the most standard track of the Tokaido Shinkansen shown in Table 1 are described as an example.

Table 1: Major specifications of tested track.

item	specification
Rail	JIS 60kg
rail pad	50MN/m (t=12mm)
sleeper spacing	581mm
ballast depth	30cm (below sleeper)

Track and rail surface irregularities which are measured by inspection cars are inputted as the forced displacement for hydraulic actuators. The included wavelength is ranges form 0.5m to 150m. The assuming train speed is 270km/h, and the unsprung-mass of the TRADYS weighs 1.8ton.

Experimental data are shown in Fig. 5. Discussions about them are described in comparison to numerical simulation results in the next section.

6 Results of numerical simulations

Results of numerical simulations are shown in Fig. 6. The forced displacement of the numerical simulation (Fig. 6(a)) is calculated as an output data. It matches very well with the input data shown in Fig. 5(a). It means that the numerical stable solution is obtained here as described in the Section 4.



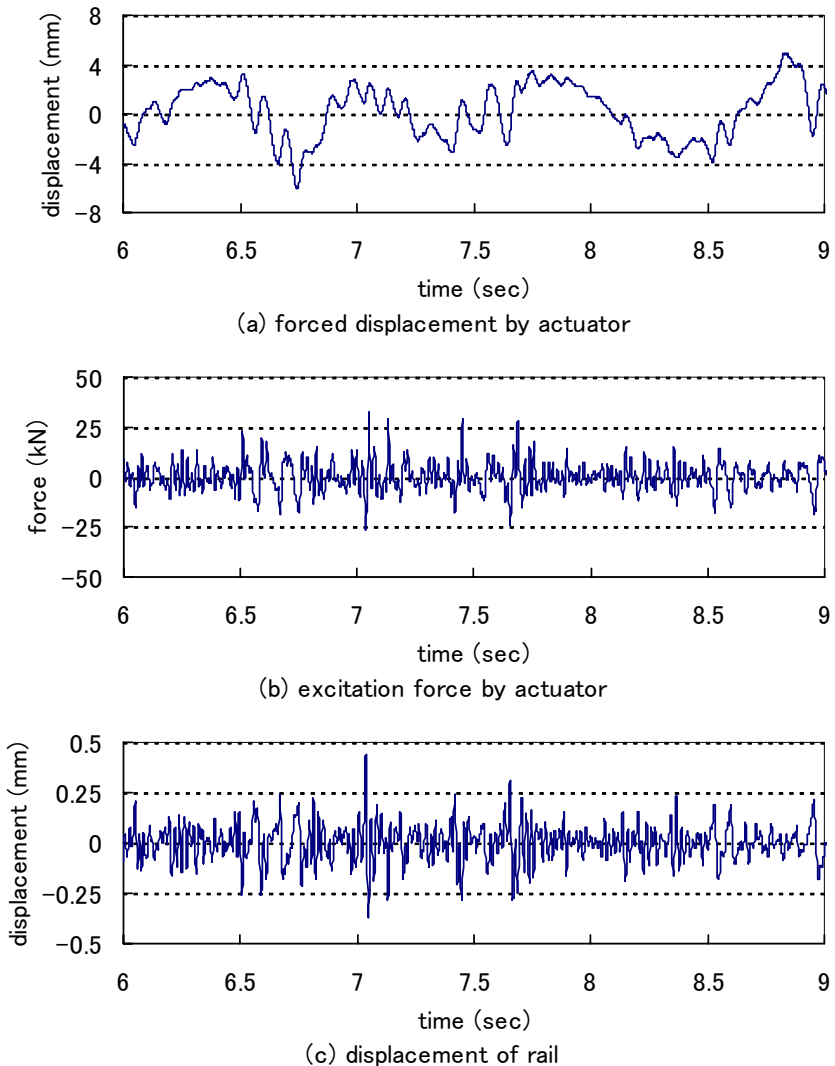


Figure 5: Experimental data in the TRADYS.

In comparison with the input data shown in Fig. 6(a) which corresponds to the track irregularity, higher frequency waves become dominant in the output data shown in Fig. 6(b) which corresponds to the dynamic wheel load. This can be similarly described in the case of TRADYS, and then Figs. 5(b) and 6(b) coincide with each other relatively well. Similar coincidence is observed for Figs. 5(c) and 6(c) which correspond to rail displacement.

The Spectral densities of TRADYS and simulation are shown in Fig. 7. The forced displacement by the actuator of each case matches with each other



significantly (Fig. 7(a)). The spectrums lower than 20Hz are resulted by track vertical irregularities which are caused by unequal settlement of sleepers, whereas those of higher than 20Hz are resulted by very small unevenness of rail surface which are stamped unavoidably by the production process (Miwa and Nakamura [1]). The dip near around 20Hz is resulted by measuring amplitude characteristics of the track inspection vehicles of two kinds, therefore it is not an essentially significant phenomenon.

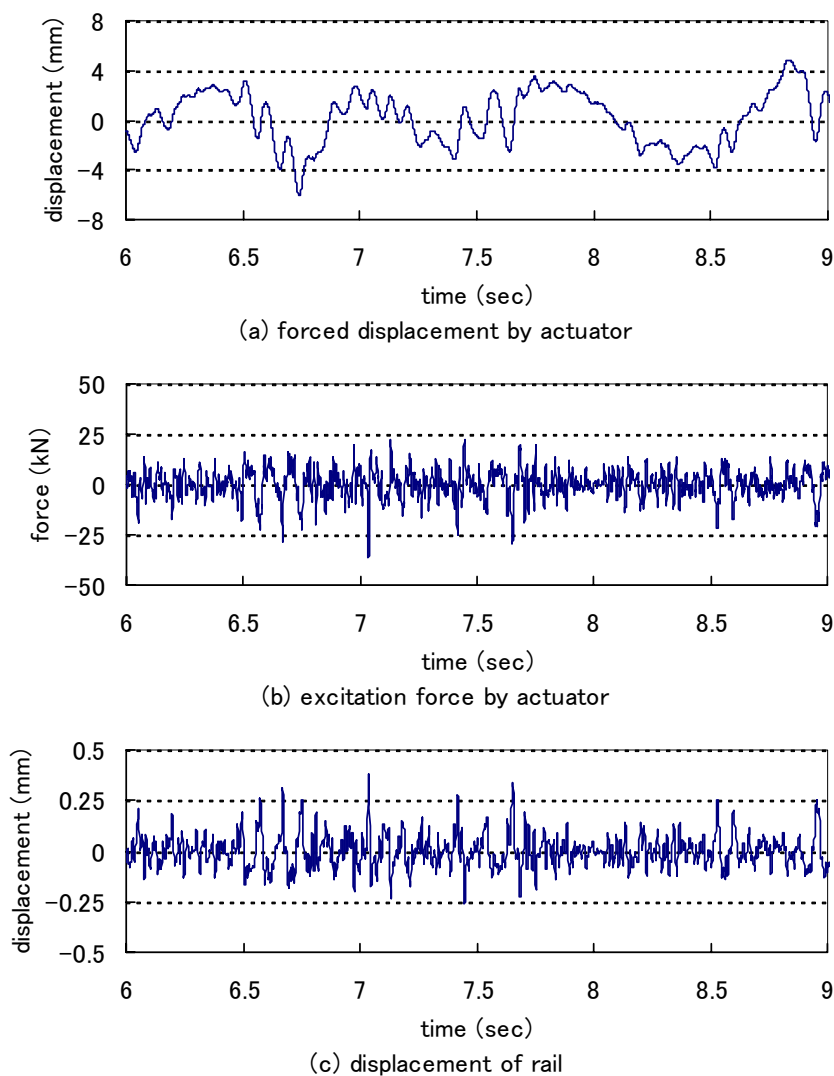
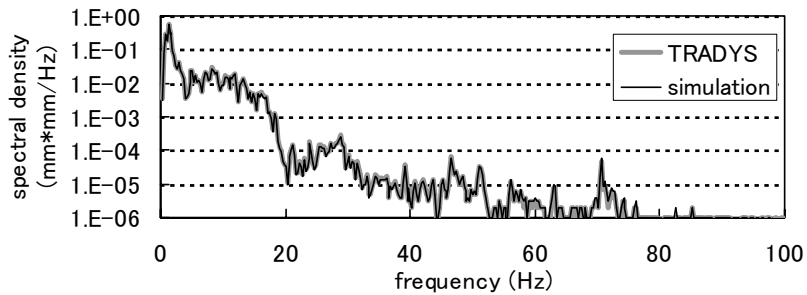
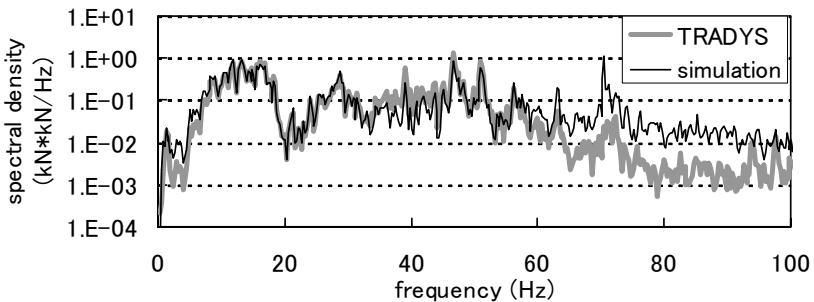


Figure 6: Numerical simulation data.

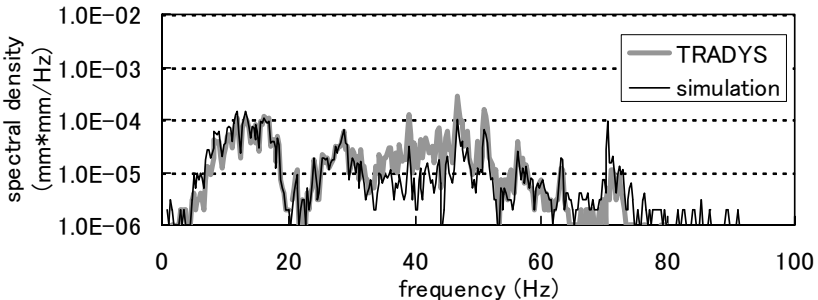




(a) forced displacement by actuator



(b) excitation force by actuator



(c) displacement of rail

Figure 7: Power spectra of the TRADYS and simulation data.

It is observed that the excitation force by the actuator in Fig. 7(b) is amplified mainly from 50Hz to 60Hz based on the input shown in Fig. 7(a). The reason is that this phenomenon is affected by the transfer function (Miwa [1]) of mass-spring system which consists of vehicle unsprung-mass and track stiffness. The track in the Tokaido Shinkansen Line has appropriate damping factor. Therefore the transfer function forms a gently sloping convex, not a steep peak called as resonance. Some peaks at 47Hz, 51Hz, 56Hz, etc, depend on the characteristics of the input data in Fig. 7(a) which has these peaks at the same frequencies in origin. In comparison between TRADYS and simulation in Fig.



7(b), they matches well in the frequency range of lower than 30Hz. Slight difference between them is seen in the frequency range between 30Hz and 50Hz, and significant deviations occur in higher than 60Hz. Referring to accuracy, higher than 60Hz, the simulation is considered to be more faithful than the TRADYS judging by the field test result (Miwa [1]). The reason why the TRADYS value is lower is that the response of load sensor comes short in higher frequency. So we are planning an improvement in the experimental method.

In the Rail displacement in Fig. 7(c), they match well with each other in lower than 30Hz similar with the case in Fig. 7(b). The difference which is seen in higher than 60Hz is smaller than that of the actuator force.

7 Conclusions

It has become possible for us to estimate complicated parameters in track dynamic model with efficiently combining two tools; (1) the TRADYS test facility which has a special vibration exciter with the same mass-spring systems as the real high-speed railway vehicle, and (2) the computer simulation in which the TRADYS on track is modeled. Moreover, adding field tests with real vehicle and track to them, we can establish powerful approaches to solve the vertical vehicle-track dynamic interaction problem.

We are now planning to extend the simulation including soil foundation model. Special roadbed for these experiments has been equipped in our research complex.

References

- [1] Miwa, M. & Nakamura, S., High speed running test and the new testing equipment for wheel/rail dynamic interaction, *Proc. of Int. Symposium on Speed-up and Service Technology for Railway and Maglev Systems (STECH'03)*, JSME, pp.489-493, 2003.
- [2] Yoshimura, A. & Ishida, M., A study on parameter estimation methods in the railway vehicle and track dynamic model, *Computers in Railways VIII (COMPRAIL2002)*, pp.201-210, 2002.
- [3] Bathe, K.J., *Finite Element Procedures*, Prentice-Hall, Inc., 1996.
- [4] Masuda, Y., Technology development at JR Central. *International Schienen der Welt*, May 2003, pp.10-17, 2003.



The study of design on the urban track transportation vibration prevention: floating slab track in the bored tunnel at Taipei

Sy. Chang¹, K. Y. Chang², K. H. Cheng³ & R. J. T. Chen⁴

¹*China Engineering Consultants, Inc. Taipei, Taiwan, Republic of China*

²*Technology Dept. of Civil Engineering, National Taipei University, Taipei, Taiwan, Republic of China*

³*Central District Project Office of DORTS, TCG, Taiwan, Republic of China*

⁴*Special Advisor of Bureau of HSR, MOTC, Taiwan, Republic of China*

Abstract

Population concentration and the resulting traffic jams are a common phenomenon in modern society. Such problems are getting worse as time goes on. Although the highway area increases every year, traffic jams still can't be solved effectively. So the popularity of land transportation is decreasing in favour of track transportation. Track transportation has many advantages, but the noise and vibration induced from the wheel/rail contact are sensitive to the neighboring residents along the track route.

There are many complaints about noise and vibration from the residents along the Taipei MRT (TRTS) line where the TRTS tunnel is underneath the building, so TRTS decided to use one of best solutions to vibration prevention available to the track engineer – Floating Slab Track (FST).

In this paper, the overall issues of the FST based on consideration of the local area in the design stage were studied, including the mode of structure analysis, the vibration isolating component, material, localized geotechnical parameters, rail fastener, elastic coefficient, etc. We hope this paper can share our practical experience and the results of the study for the vibration prevention in track at the urban transportation for environmental pollution prevention, to other railway systems or to academia in the environment field and railway field.

Keywords : *vibration, localize, Floating Slab Track, vibration prevention, elastic coefficient.*



1 Foreword

Population concentration and the resulting traffics jam are a common phenomenon in modern Taiwanese society. Such problems are getting worse as time goes by. Although the roadway area increases every year, traffic jams still can't be solved effectively. So the popularity of land transportation is decreasing in favour of track transportation. They have a many complains about the noise and the vibration from the Taipei MRT (TRTS) operation at the Chungho line, the noise and vibration problems produced by the train/track interface, due to the wheel/track contact, which generates friction and impact. This vibration effectively propagates through the ground to the neighbouring buildings along the track route, which will transmit noise and damage to sensitive equipment. The modern railway will solve noise and vibration problems by track/ EMU/ geology/ structure. The track has elastic rail fasteners, elastic rail boots, elastic trackbed (floating slab, LVT, tie with elastic boots). The Xinzhung/Luzhou new line has room to pass under the existing buildings on the line and will use the floating slab track to solve the vibration problems. This paper will provide some suggestions for the design details of the floating slab to be used in Taiwan area, and provide references and data for the design and selection of the floating slab in this area.

2 The source of track vibration

The vibration of track was cause by the following reasons when the railroad train is running on the track (see the figure 1).

- (1) Wheel roughness
They have difference maintenance values of wheel truing in different track transportation systems, so the track design has to following the maintenance figure of the wheel truing of the Taipei Rapid Transit Company.
- (2) Rail corrugation
The characteristic of the rail corrugation / the rail wear wave-pattern will depend on the local track transportation system, choosing the wrong factor will cause a more serious vibration problem, so the track design has to follow the maintenance figure of the rail grinding of the Taipei Rapid Transit Company.
- (3) Wheel and rail grinding abrasion metal chipping.
- (4) Discontinuous rail
Taipei Rapid Transit system uses the non-rail expansion join of the continuous welding rail system, which has less vibration problems. The rail insulated joint shall away from the gap between two floating slab, in case of the abnormal loading in the rail insulated joint caused by the different deformation between two track slabs.
- (5) Roughness of the track supporting structure



The Taipei Rapid Transit system was constructed using a concrete track plinth on the mainline and concrete track slab in special track areas, so it should have high accuracy of track construction, it should consider the uniform of longitudinal stiffness in the transition section of floating slab connected to the main line track, the re-evaluation of the elastic coefficient of rail fastener base plate, and transition design of the floating slab.

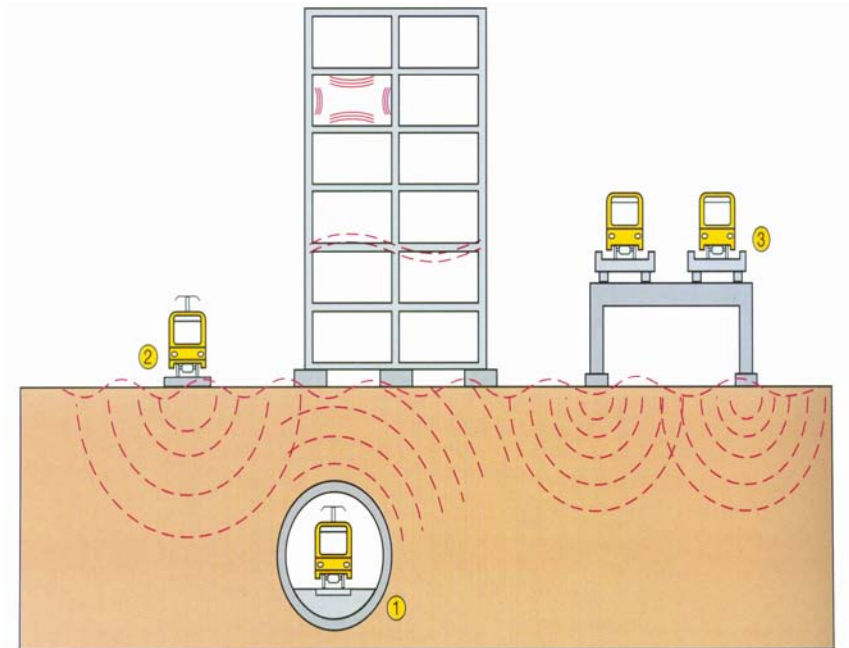


Figure 1: Track ground-borne Vibration Propagation Route (GERB).

The above mention reasons will cause the track vibration, and this source of the vibration will use the following way to the ground.

- (1) Track system
The track system was including the rail and rail fastener and concrete trackbed.
- (2) Tunnel structure
The tunnel structure types have cut and cover and bored tunnel.
- (3) Geotechnical conditions
The soil type and the ground born-vibration deduction factor are the main key issues of the floating slab design, so the track contractor should have a site investigation and measurement along the line to define the ground born vibration deduction factor.



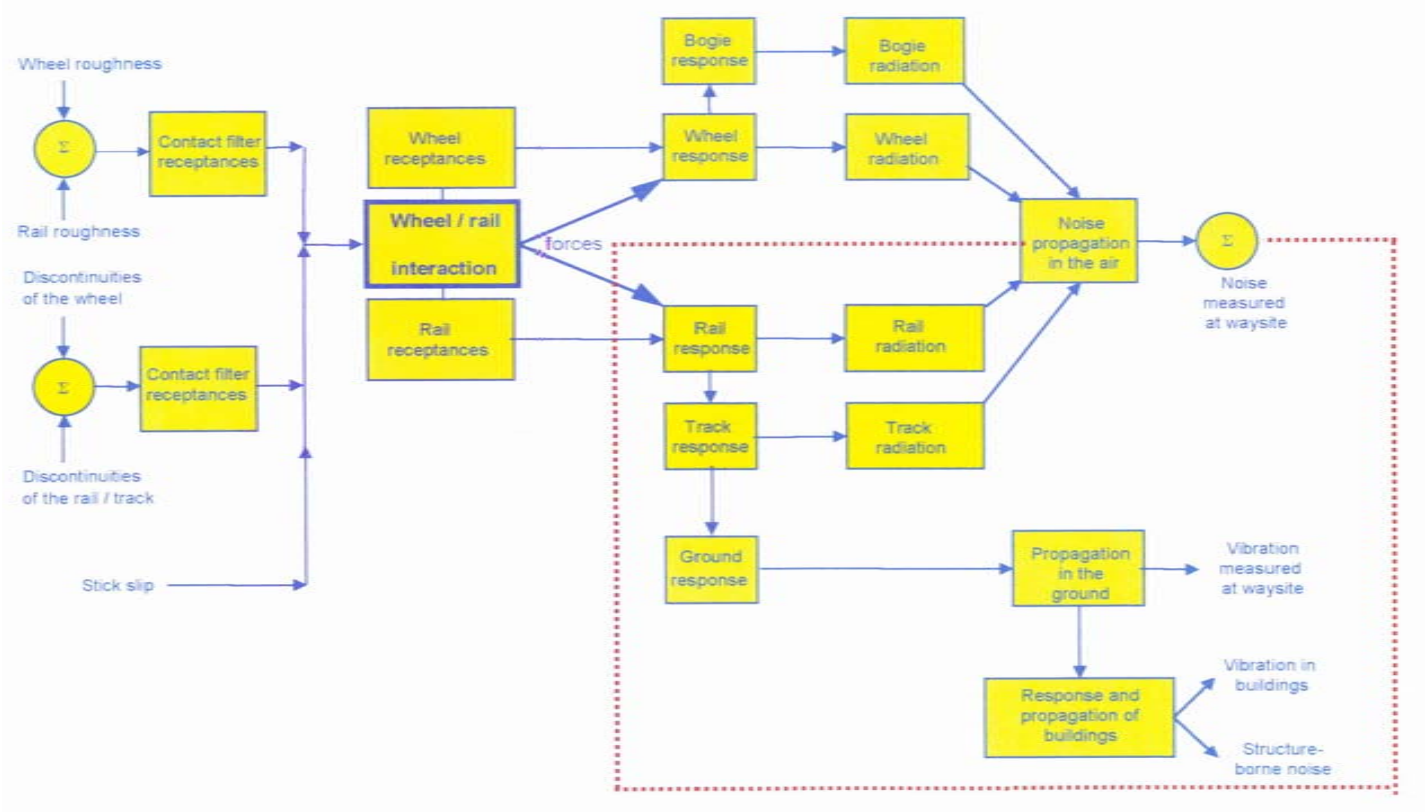


Figure 2: Agglomerate of vibration [CDM].

- (4) The resident structure foundation inside the structure
Different types of structure will directly affect the vibration measuring value inside the adjacent buildings along the line; it should be investigated carefully (the investigation should include structural stiffness, resonance frequency, number of floors, etc.) of the adjacent building along the line. These kinds of vibration maybe affect the living quality of resident and some equipment operation, but it would not to damage the adjacent building along the line.

3 The vibration reduction of the floating slab

The floating slab under the track and the EMU was the source of vibration of track transportation system. The floating slab design was based on the Spring-Mass System. The floating slab design needs to go through overall systematic consideration. The design parameters that need to be considered include: the mass and stiffness values of each components of the train, mass and elastic coefficient / damping ratio values of each track component, structural types and geotechnical parameters, the wheel / rail interface during contact and other practical operational considerations.

The floating slab used a concrete track slab with an elastic support to provide vibration attenuation and energy absorption, and the heavier the mass of track slab will have a better result. However, a larger track slab will need more room in the civil structure, increasing the initial cost of the civil structure and at the same time increase the initial cost of the floating slab. Also the mass of the track slab must be suitable for the elastic support: the softer the support the better the vibration attenuation, but a soft support will lead to rail deformation that will make the track unstable, so an appropriate elastic coefficient should be selected.

Both track slab and isolating / deduction of vibration system should have a overall consideration of the natural frequency of the floating slab, to avoid resonance between the car body and track and civil structure and soil propagation and adjacent buildings (the soil propagation frequency rang was within 25-150Hz, each parts of car body was within 1-60Hz).

To avoid damage to the floating slab as trains run on the track (natural frequencies in the range of 5-150Hz), one has to perform a static and dynamic analysis in the track design: the vibration analysis of the track has a one car length vibration model (see figure 3), weight of car body, M_c , primary suspension system elastic coefficient, K_{s1} , damping ratio, C_{s1} , mass of bogie, M_t , secondary suspension system elastic coefficient, K_{s2} , damping ratio, C_{s2} , wheel load, M_w , and dynamic loading transfer to the track $P_1(t)$, $P_2(t)$, $P_3(t)$, $P_4(t)$.

The early stage of conceptual design uses a simple model to calculate the natural frequency of each element of the track structure away from the natural frequency of car and soil.

We have studied the Taipei mass rapid transit Xinzhung/Luzhou line and perform a simple model calculation, and gained the following results



Nature frequency W_n (Hz)

Case 1

Rail base plate mass=2144kg
elastic coefficient=25kN/mm*8=200kN/mm
bearing pad / floating slab mass=5163kg
elastic coefficient=8kN/mm*8=64kN/mm

Case 2

Rail base plate mass=2900kg
elastic coefficient=25kN/mm*20=500kN/mm
bearing pad / floating slab mass=26710kg
elastic coefficient=8kN/mm*20=160kN/mm

Case 3

Rail base plate Mass=2013kg
elastic coefficient=25kN/mm*8=200kN/mm
Bearing pad / floating slab Mass=5032kg
elastic coefficient=10kN/mm*8=80kN/mm

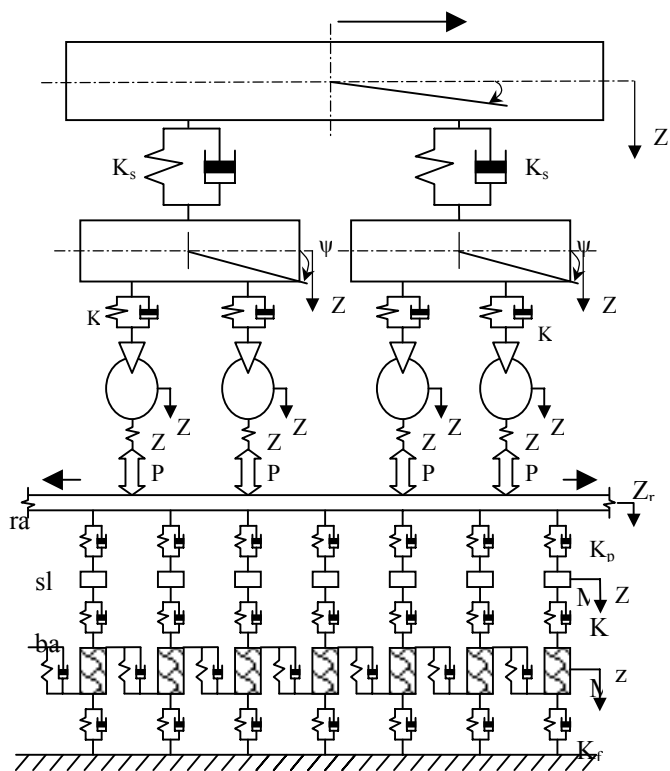


Figure 3: Vertical integration system of track.



Table 1.

	Mass (kg)	Elastic coefficient (kN/mm)	Natural frequency (Hz)
Bearing pad / FST (M1, K1)	5163	64	15.052
Rail base plate (M2, K2)	2144	200	58.853
Primary suspension system (M3, K3)	5847	2.67	3.692
Secondary suspension system (M4, K4)	12250	0.57	0.972

Table 2.

	Mass (kg)	Elastic coefficient (kN/mm)	Natural frequency (Hz)
Bearing pad / FST (M1, K1)	26710	160	11.793
Rail base plate (M2, K2)	2900	500	69.841
Primary suspension system (M3, K3)	5847	2.67	3.741
Secondary suspension system (M4, K4)	12250	0.57	0.976

Table 3.

	Mass (kg)	Elastic coefficient (kN/mm)	Natural frequency (Hz)
Bearing pad / FST (M1, K1)	5,032	80	17.014
Rail base plate (M2, K2)	2013	200	60.587
Primary suspension system (M3, K3)	5847	2.67	3.705
Secondary suspension system (M4, K4)	12250	0.57	0.973



4 Conclusions and recommendations

For the study of floating slabs in Taiwan we know that a full picture analysis in the design and construction stage has to be produced, but the different track transportation systems have different characteristics, and different problems that need to be solved, so we summarise some recommendations as follows:

- (1) It has to measure the natural frequency of the car body.
- (2) The optimization design of the floating slab should follow the results from track and structure analysis.
- (3) The detail investigation of types of adjacent buildings and detailed analysis of the structure characteristics to avoid vibrational resonance.
- (4) It should carefully analyse the impact factor of ground-borne vibration, especially for the soil: this information should be established by the site measurement and the construction data.
- (5) To cooperate with the maintenance plan of the operation unit that will get the optimal design of floating slab.

For high speed railways, mass rapid transit system, traditional railway systems or light rail system the floating slab is still be the best and most efficacious solution, but its construction cost will be 3 to 10 times the non-ballasted track, so most systems will re-evaluate its necessity before using it, and its real vibration deduction will depend on the accurate analysis, and adequate design and construction. This paper collected local design factors for the future development of a Taiwanese floating-slab track, and we hope this paper can serve as a useful tool for the future use of floating slab tracks in Taiwan and share our experience worldwide.

References

- [1] Kuo H. Cheng et al, The Study on The Floating Slab for Reduce The Noise and Vibration, 2002 World Metro Symposium, Taipei-Taipei City Government, 25-27 April 2002.
- [2] Sy Chang et al, Trackwork Innovation-Non-Ballasted Track In Taiwan, Eighth International Conference on Computers in Railways -COMPRAIL VIII, Wessex Institute of Technology, 12-14 June 2002.
- [3] The study on the Floating Slab – an alternative to reduce the noise and vibration level of non-ballast track.



Section 16

Freight

This page intentionally left blank

Effect of the distribution of the arrivals and of the intermodal unit sizes on the transit time through freight terminals

G. Malavasi, A. Quattrini & S. Ricci

D.I.T.S. Transport Area, University of Rome "La Sapienza", Italy

Abstract

The railway freight terminals play a key role within the multimodal logistic chain. Therefore, the transit time through these terminals represents one of the most relevant terminal performances and at the same time a key component of the freight transport generalized cost, so that its quantitative analysis is a strategic activity, both in the terminal planning (for the existing and for the future terminal) and in the logistic chain organization. The transit time is composed of deterministic and stochastic components, which increases significantly the problem complexity. The authors developed an original model based on queuing theory allowing the calculation of the transit time for a large variety of terminals. Some pilot applications of the model showed that the arrival distributions of both classes of vehicles are critical inputs for the effectiveness of the results. In this framework the present paper focuses on the theoretical study of these distributions including an extended validation campaign based on data available for an Italian case study (intermodal terminal of Pomezia near Rome). Another relevant aspect, affecting the quality of the model results, is the presence in the terminal of different sizes of intermodal units. In the paper a methodology to take into account these size differences effectively is also explained.

Keywords: intermodal, freight, terminal, transport unit, arrival distribution.

1 Introduction

The authors developed an original model [2–4] based on the queuing theory allowing the calculation of the transit time (*TTR*) of the freight transport units through the intermodal terminals.



In this framework a refining of some key parameters of such model has been performed. After a synthetic description of the model structure, a methodological approach based on real collected data taking into account the influence of the intermodal unit size is exposed.

2 Methodological approach

The first step is to adopt a common transit time definition. Here is adopted the following one: “time period from the arrival of the single freight unit to the terminal gate from an external transport infrastructure to its exit from the terminal towards a different transport infrastructure”.

The second step is the formalization of the model finalised to the determination of the transit time. With this purpose the following general and, nevertheless, simple formalization has been selected:

$$TTR = TE(C, O) + TI(T, D, R) \quad (1)$$

where:

- TE depends upon external infrastructures and transport services, to be formalized in two sets of parameters and constraints:
 1. carrying capacity C (e.g. railway line bottlenecks, etc.)
 2. operation planning O (e.g. timetable structures of the transport services arriving to the terminal and departing from it, etc.);
- TI depends upon technologies, dimensions and operational rules in the terminal, to be formalized in three sets of parameters and constraints:
 1. terminal planning parameters T (e.g. check-in and transfer technology);
 2. terminal dimensions D (e.g. distances between gates and transfer area, number of tracks, etc.);
 3. operational constraints, rules and regulations R (e.g. speed limits, maximum loading weights, etc.).

In order to define a common sequence of operations to be performed within the terminals, the single activities have been analyzed into details.

For each activity have been identified:

- an Operational Phase (OP) and a previous Waiting Phase (WP);
- the corresponding durations, Operation Time (TO) and Waiting Time (TW).

The following list shows the single phases which have been identified for the most general cases:

1. Waiting before entering the terminal + Entering movement;
2. Waiting before check-in + Check-in operations;
3. Waiting before the first units transfer + First units transfer;
4. Waiting before the second unit transfer + Second units transfer;
5. Waiting before check-out + Check-out;
6. Waiting before exiting the terminal + Exiting movement.

In each generic terminal two different classes of entering and exiting vehicles (V' and V'') may be identified.



In the most general case V' and V'' may allow the transport of a very different amount (NU' and NU'') of freight units. For instance in the land freight interchange is $NU' \text{ (truck)} < NU'' \text{ (train)}$. Accordingly, in each terminal opposite flows of freight units entering on V' or V'' and exiting correspondingly on V'' or V' may be identified.

3 Transit time calculation by model

In figure 1 the duration of the single phases and the mean total transit times for the freight units running in both the directions (TTR' and TTR'') are represented. In the scheme of this figure a generic freight unit entering on a vehicle V' runs on it towards the transfer area and, after the stocking phase, proceeds on the vehicle V'' . Therefore, in the most general case, the unit is transferred twice: from V' to the stocking area and from there to V'' . Whenever a direct transfer from V' to V'' is possible, the second transfer phase ($OP4$) is not performed and correspondingly times ($TO4$ and $TW4$) are equal to 0.

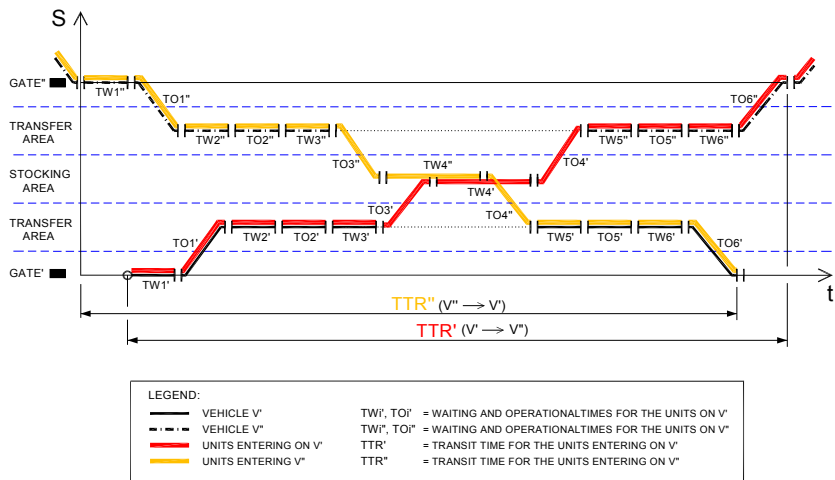


Figure 1: Single phases duration and total transit time in a generic terminal.

According to the unit flows within the plant represented in figure 1, the transit time may be formalized as follows:

$$TTR' = \sum_{i=1}^4 TW'_i + \sum_{i=1}^4 TO'_i + \sum_{i=5}^6 TW''_i + \sum_{i=5}^6 TO''_i \quad (2)$$

$$TTR'' = \sum_{i=1}^4 TW''_i + \sum_{i=1}^4 TO''_i + \sum_{i=5}^6 TW'_i + \sum_{i=5}^6 TO'_i \quad (3)$$



Eqn. (2) represents the TTR for the freight units entering on V' and exiting on V'' ; eqn. (3) represents the TTR for the opposite flow (units entering on V'' and exiting on V').

4 Validation campaign

4.1 Data collection

The above-described generalized model has been applied to an Italian case study: the Pomezia freight intermodal terminal, located about 30 km south of Rome.

The authors collected a large quantity of data related to the transport units entering/exiting in/from the terminal on the two transport modalities. These data allowed the calibration of some key parameters to optimize the model formulation; moreover the study of the unit arrival/departure distribution is necessary to determine important input of the model (e.g. the mean time between 2 arriving trucks/trains).

In this context the paper focuses on the transfer of the units entering from the road side (by trucks – V') and exiting on the railway side (by trains – V'') of the terminal for the TTR' value calculation.

The following sections show the analysis performed on the collected data and its results.

4.2 Arrival/departure distributions

Figure 2 shows a graphic view of the collected data: the x -axis represents the terminal functioning period, the y -axis represents the number of units arriving to the terminal by truck (various colours correspond to different investigated days). Similar investigations have been carried out also for unit departing by truck.

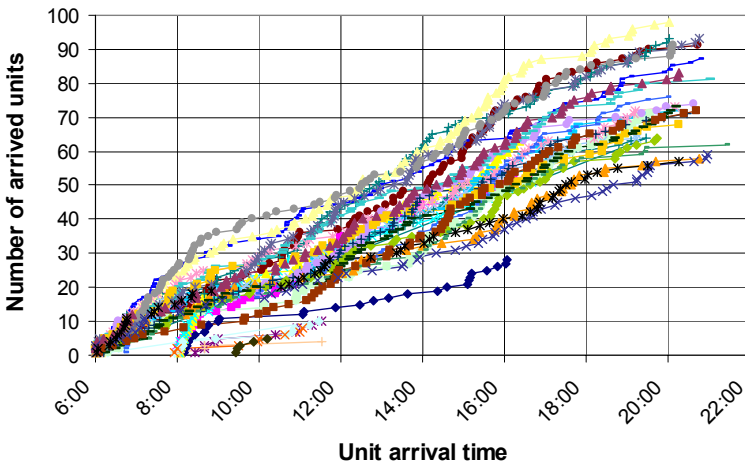


Figure 2: Units arriving by truck to the Pomezia terminal – collected data.



The collected data have been elaborated to calculate the typical operational range, in which the most part of the real cases are included. The results of such elaboration are shown in figure 3.

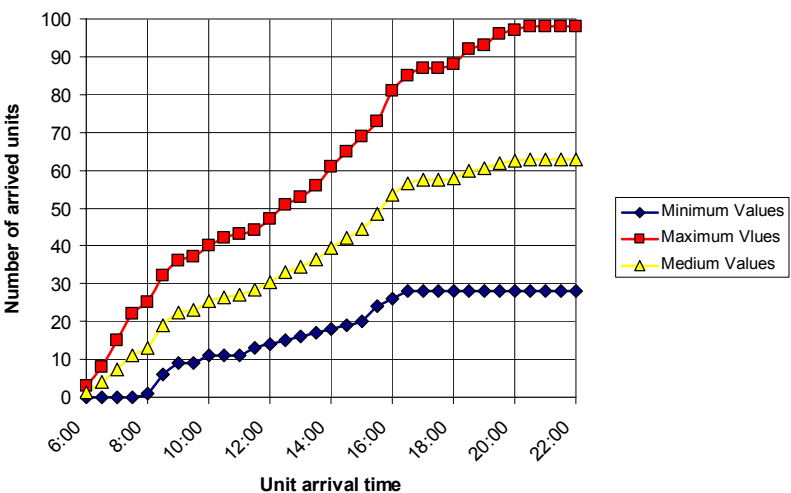


Figure 3: Distribution of units arriving on trucks – Pomezia Terminal.

The two curves of the figure have been obtained considering extreme and medium values occurring during the reference period.

- Similar analyses have been performed for the data collected on the train side. The results of such operations are shown in figure 4.

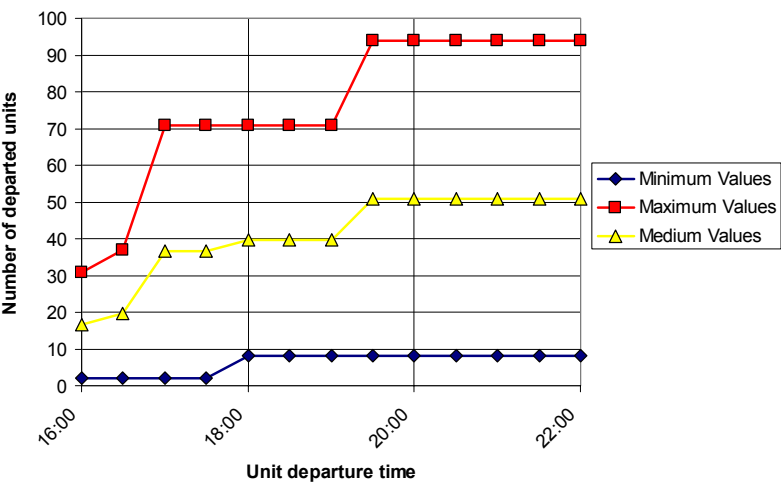


Figure 4: Distribution of units departing by train – Pomezia terminal.



In the figure 4 case, the shape of the curves is quite different from the previous figure 3. This is due to the fact that the number of trains departing from the terminal is largely lower than the number of arriving trucks.

Finally, the procedure shown in figure 5 has allowed the TTR' evaluation on the basis of the above-reported results (medium values graphically displayed by the yellow curves in figures 3 and 4).

The results obtained for the Pomezia terminal are the following:

- $TTR'(min) = 3h\ 30'$;
- $TTR'(max) = 8h\ 00'$;
- $TTR'(med) = 5h\ 45'$.

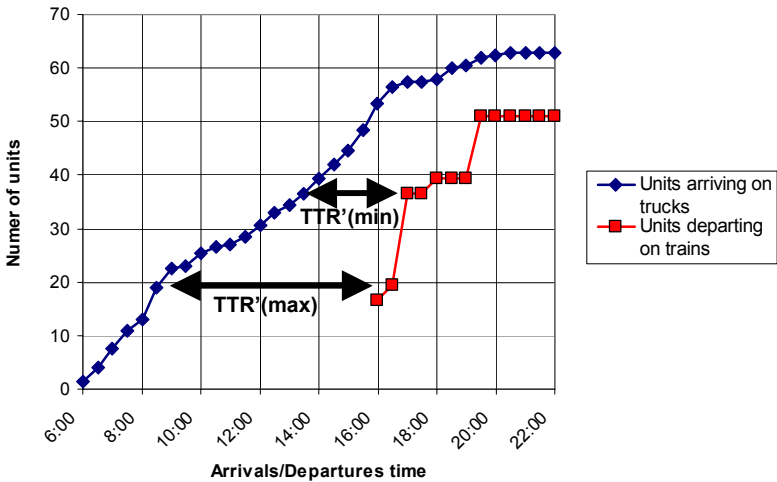


Figure 5: Graphic representation of $TTR'(max)$ and $TTR'(min)$ – Pomezia Terminal.

4.3 Unit size variability

Another aspect affecting the sensitivity of the model results is the presence in the terminal of different sized transport units; in fact with the same quantity of Twenty feet Equivalent Unit (TEU) is possible to handle a different number of intermodal transport units.

The EIA (European Intermodal Association), on the basis of extended investigations, suggests the following conventional equivalence between TEU and ITU :

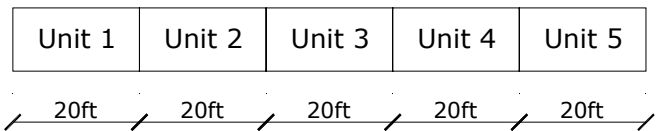
$$1,4 \cdot ITU = 2,3 \cdot TEU \Rightarrow 1 \cdot ITU = 1,6 \cdot TEU \quad (4)$$

On this basis, figure 6 shows two typical situations that can occur in the terminal: case b (corresponding to the EIA conventional value) is easier to work because the transshipping device can transfer the same freight quantity with less handled units than in case a ($100\%TEU$).



Accordingly to these observations, the *TTR* is expected to decrease when the *TEU / ITU* ratio increases.

Case a: 5 Units = 5 T.E.U.



Case b: 3 Units = 5 T.E.U.

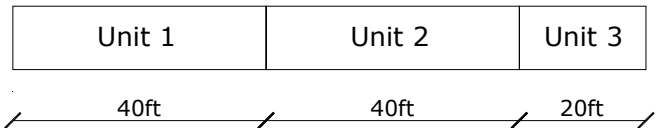


Figure 6: *TEU – ITU* correspondence.

5 Case study application

Figure 7 is a graphic view of the model application results to Pomezia terminal on the basis of the collected data.

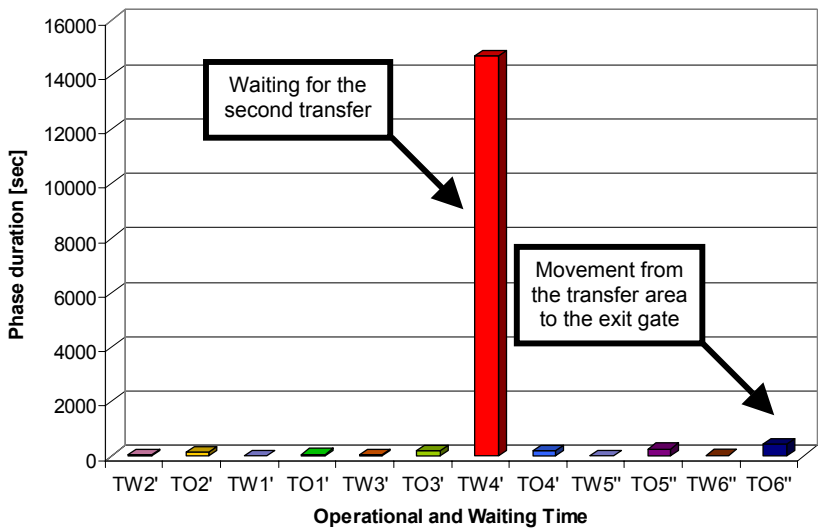


Figure 7: Model application results – Pomezia Terminal.

The analysis of the numerical values of the single time components calculated by the model leads to the following considerations:

- calculated TTR' value is about $4h\ 30'$, so it provides a good representation of the real Pomezia situation shown in figure 5;
- for the units entering by trucks, the waiting for the second transfer in the stocking area ($TW4'$) is largely the most important period within the terminal (about 91,6% of the global transit time);
- other important time period is due to the transport unit movement from the transfer area to the exit gate (it depends on the long distance that the unit must go through);
- in case of direct transfer from the first to the second vehicle (without stocking phase: $TW4' = TO4' = 0$) the global transit time is strongly reduced;
- the internal parameters have a very limited influence on the global transit time;
- the waiting times are largely higher (92% of TTR') than the operational ones.

Figure 8 shows the TTR' sensitivity to the TEU/ITU ratio variation. The figure indicates an appropriate reduction of the TTR' values when the ratio increases (less handled transport units with the same freight quantity).

Anyhow, on the basis of the model results, the TTR' decreases extremely slowly; this is due to the external parameters that have a stronger effect than the internal ones on the global transit time value. Therefore, the operational activities of the terminal weakly affects the TTR' value.

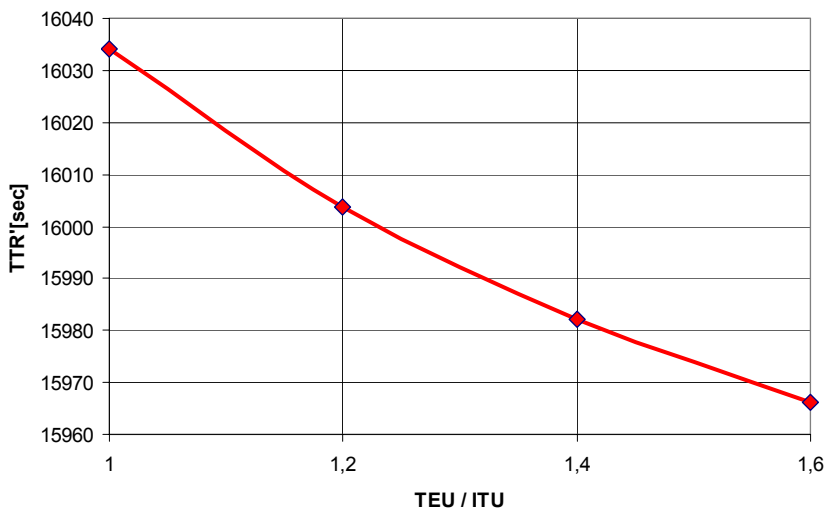


Figure 8: TTR' sensitivity analysis to the TEU / ITU ratio variation.

6 Conclusions and future research developments

The developed model is characterized by wide generality and applicability to different terminal typologies, lay-outs, dimensions and transfer technologies.

The approach allows to highlight contributions and weights of the various activities and phases of the freight unit transit through the terminal by distinguishing operational and waiting periods, whose duration depends not only upon internal performances (technologies, dimensions and operational rules) but also upon external parameters and constraints (time distribution of vehicles arrivals and departures).

Therefore, this model can be widely helpful in the following application fields:

- Design of new terminals in the following design phases:
 1. New terminal dislocations
 2. Pre-dimensional study
 3. Functions dislocations
 4. Transshipping device choice
- Optimization of existing terminals by identifying the existing bottlenecks.

At present, the results of the application to some case studies show that the amount of time dedicated to the activities only depending upon internal parameters is generally very limited (about 5% in the land freight interchange case studies).

According to these results and to the present state of development, the research is going to be further developed in the following directions:

- consolidation and tuning of the methodological approach by means of a wider set of case studies applications based on the comparison with experimental data;
- model explication finalised to the systematic identification of internal and external carrying capacity bottlenecks;
- integration of the proposed model within a wider procedure which leads to the calculation of the total transport time from the origin to the final destination of the goods.

The performed analyses highlight that the application of the model to terminals which handle larger quantities of transport units is extremely important for a further model validation. These additional applications are a basic element to better evaluate the model results particularly in case of units entering by train (smaller quantity of events). Consequently, the authors are planning to extend the model application to a greater number of large and much exploited terminals.

References

- [1] Arnold, D., Rall, B., *Analyse des Lkw-Ankunftsverhaltens in Terminals des Kombinierten Verkehrs*, Internationales Verkehrswesen 6, 1998.
- [2] Malavasi, G., Ricci, S., A synthetic model for the evaluation of sea-side port terminal capacity. *Proc. of the International Workshop of the Special*



- Interest Group on Maritime Transport and Ports (SIG-2)*, World Conference on Transport Research Society, Genoa 2000.
- [3] Malavasi, G., Ricci, S., Generalized model for the performance evaluation of different railway freight terminals. *Proc. of the 9th International Symposium Zel 2002 – Railways on the Edge of the 3rd Millennium*, University of Zilina – CETRA, Zilina 2002.
 - [4] Malavasi, G., Quattrini, A., Ricci, S. Personal communication, 18 november 2005, Urbanpromo 2005 Scientific Seminar - “Interporti, scali merci, centri logistici e sviluppo competitivo”, Venice, IT.
 - [5] Malavasi, G., Zanolin, S., Formazione di treni merci e transito negli impianti di smistamento, *Ingegneria Ferroviaria* 8, 1997.
 - [6] RAIL Cargo Austria (Company). Cargo terminal saves space. *Railway Gazette International Vol. 159 Issue 3*, March 2003.
 - [7] Vandever, D., Intermodal rail facility design for the next century. *Transportation Research Circular 459*, June 1996.
 - [8] Verdon, L., Rousse, R., A menagement des gares terminales pour containers et equipements de manutention, *Rail International* 3, 1971.
 - [9] Zimmer, R.N., Designing intermodal terminals for efficiency. *Transportation Research Circular 459*, June 1996.



COMPAT: a decision support tool for determining the necessity of rail infrastructure

M. J. Wolbers

Holland Railconsult, The Netherlands

Abstract

The functionality of a railway yard is for the largest part determined by the existing simultaneities. Simultaneities allow the simultaneous arrival, departure or drive through of trains.

At medium or large sized yards the process of manually determining all simultaneities is very time-consuming and error-prone, which limits the type and number of questions to be answered within a reasonable time, such as the question as to whether all existing infrastructure is really needed given the requirement specifications for compatible routes.

COMPAT is a new tool developed by Holland Railconsult, which uses a mathematical algorithm to determine all simultaneities on a railway yard. The results can be visualised graphically. The tool enables the user to quickly determine which points or links are and are not necessary to meet the functional requirements and to determine the added functionality when a certain point or link is added.

This means current railway yards or new designs for railway yards can be tested by COMPAT for redundant infrastructure. Therefore, COMPAT provides the railway infrastructure manager with a powerful tool to determine a cost effective layout of a railway yard. By presenting the results graphically, COMPAT contributes to the user's insight into the effects of adding or removing points and links in railway yards, thus reducing discussions among specialists.

Keywords: reachabilities, simultaneities, compatible routes, redundant infrastructure.



1 Introduction

The functionality of a railway yard is for the largest part determined by the existing simultaneities. Simultaneities allow the simultaneous arrival, departure or drive through of trains.

At medium or large sized yards the process of determining all simultaneities is very time-consuming, which limits the type and number of questions to be answered within reasonable time, such as the question whether all existing infrastructure is really needed given the required specifications for simultaneities. This is a very interesting question because of the tendency to critically assess the necessity of points.

Research revealed that there was a need within Holland Railconsult for a tool that could help answering questions regarding simultaneities [1]. Holland Railconsult has developed such a tool [2]. This paper describes the functionalities of the tool and the mathematical algorithms it uses.

2 Mathematical problem

In order to describe the mathematical problem, we need to define two concepts: reachability and simultaneity. A reachability $b(s,t)$ is defined as the ability to reach point t starting from point s over any available route in the railway yard. A simultaneity $g(b(s_1,t_1) \# b(s_2,t_2))$ is defined as the ability to reach t_1 from s_1 and t_2 from s_2 at the same time without using any common point.

Determining if a set of reachabilities form a simultaneity equals the ‘vertex-disjoint paths problem’. This problem is defined as follows [3]:

Given:

A (directed or undirected) graph $G=(V,E)$, pairs $(s_1,t_1), \dots, (s_k,t_k)$ of vertices of G

Asked:

Pairwise vertex-disjoint paths $P_1, \dots, P_k$ with P_i being an s_i - t_i -path ($i=1 \dots k$).

Here the pairs $(s_1,t_1), \dots, (s_k,t_k)$ are the reachabilities of the graph. If this has a solution then the reachabilities form a simultaneity.

This “vertex-disjoint-paths-problem” however is NP-complete even when having only two pairs [4], so there is no efficient algorithm finding all simultaneities.

The problem can also be seen as a multi-commodity network flow problem, with the reachabilities in the graph representing the commodities.

3 Determining all simultaneities

Determining all simultaneities equals the following question:

Given:

A railway yard, represented by a directed graph $G=(V,A)$ with given starting points (V_s) and ending points (V_e) .

Asked:

All simultaneities on the railway yard.



Because the number of simultaneities can go into millions on very large railway yards like Utrecht Central Station in the Netherlands, it is impossible to determine for every combination of reachabilities if it forms a simultaneity. That is why all simultaneities have been determined by enumeration.

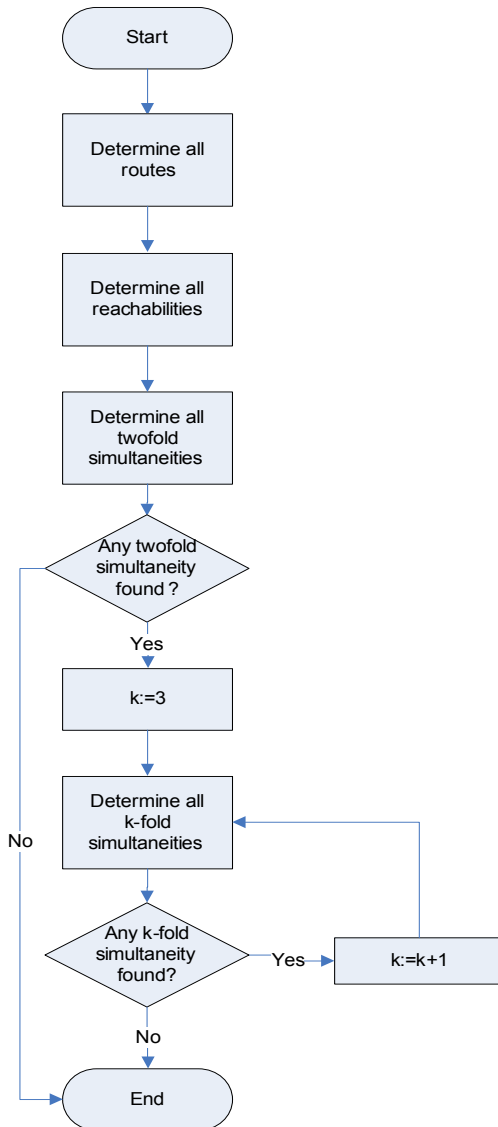


Figure 1: Determining all compatible routes.



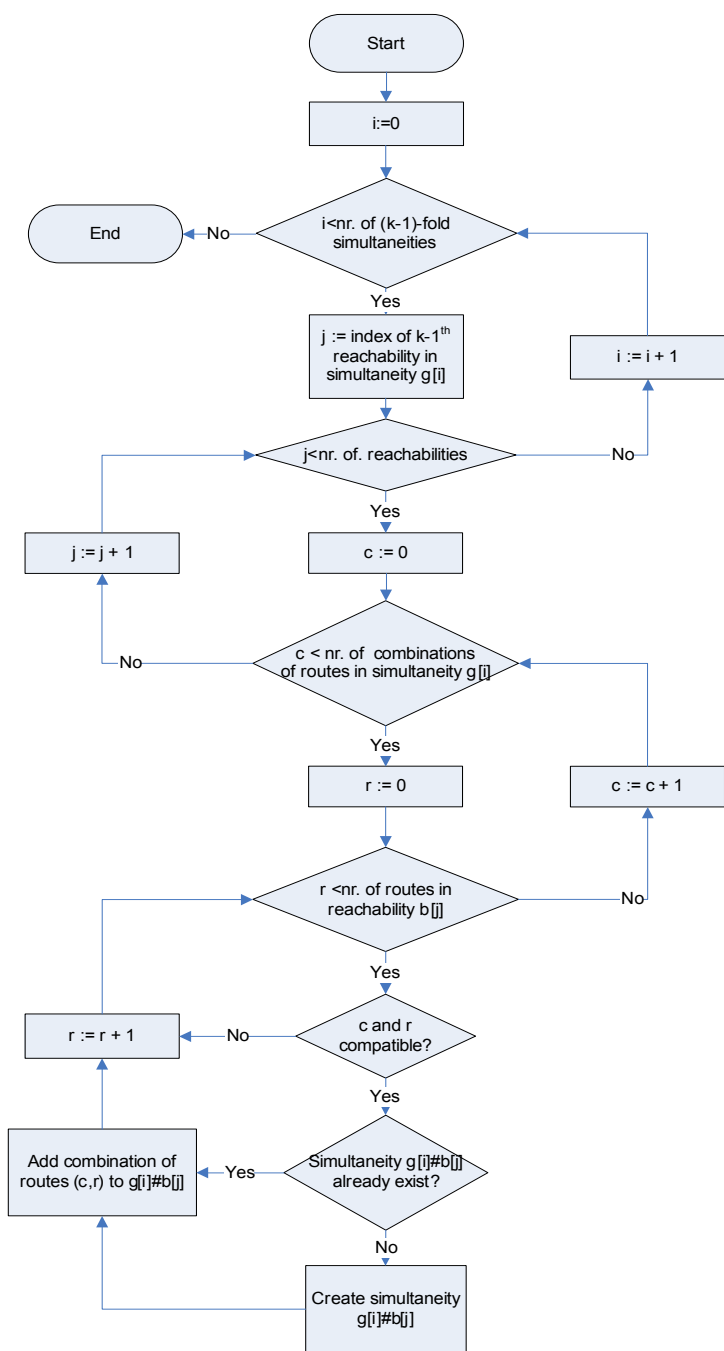


Figure 2: Determining all k-fold ($k \geq 3$) simultaneities.



To determine all simultaneities, first all routes on the railway yard from a starting point to an ending point have been determined. From these routes, all reachabilities can easily be extracted. With each reachability, all routes that start at the starting point and end at the ending point of the reachability are saved. Then all twofold simultaneities are determined from these reachabilities. If two reachabilities have both a route which is compatible, these reachabilities form a simultaneity. In each simultaneity all sets of routes that are compatible are stored. From these twofold simultaneities, the k -fold ($k \geq 3$) simultaneities etc. can be determined. This is done by trying to add one reachability to a $(k-1)$ -fold simultaneity. The process of determining all simultaneities is shown in figure 1, a closer view on determining all k -fold ($k \geq 3$) simultaneities is shown in figure 2.

4 Reducing complexity

Because the number of simultaneities on the railway yard grows exponentially when the number of starting vertices and ending vertices increases, a functionality has been added to decrease the number of routes used during the process shown in figure 1.

The complexity of the problem can clearly be reduced by removing all redundant routes from the computations. Up to 90% of all routes can be redundant in complex railway yards. A route 'r1' is redundant if it blocks exactly the same or more other routes than another route 'r2' and both routes belong to the same reachability. Put otherwise, if $BI[i]$ is the set of routes blocked by route i , then route 'r1' is redundant if there is another route 'r2' with $BI[r2]$ being a subset of $BI[r1]$.

An example of a redundant route is shown in figure 3. The upper route is redundant because it blocks the same and more other routes than the lower route.

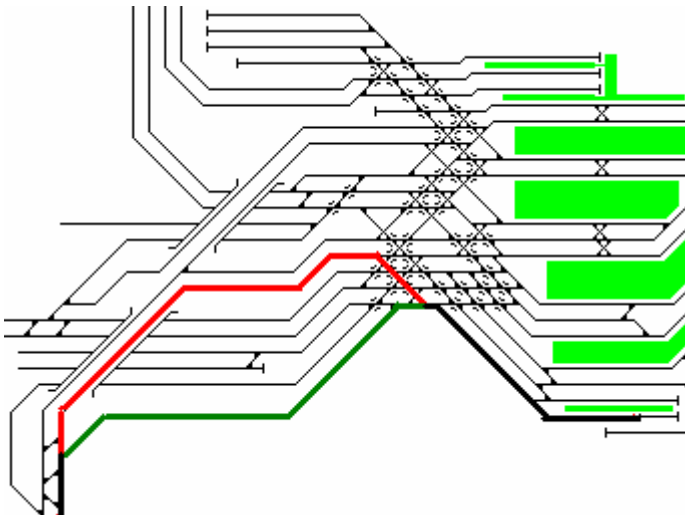


Figure 3: Example of a redundant route.



5 Main functions

5.1 Determine the difference between two alternative lay-outs

The tool allows the user to determine the difference between two alternative lay-outs, in the sense of available reachabilities and simultaneities. This means the user can get a quick view of the consequences when adding or removing a point or link, but also the consequences of more complex changes of the lay-out can be determined.

To determine the difference, the two lay-outs are merged into one new lay-out, with all points and links marked whether they exist in layout 1, 2 or both. Next, the same procedure for determining all simultaneities is followed. Every time a route has been found it is checked if that route consists of any infrastructure that was not in layout 1 and in layout 2 or vice versa. In that way it is clear if that route was in layout 1 or in layout 2 or in both.

A reachability is in layout 1 if it contains any route that is in layout 1. A simultaneity is in layout 1 if it contains a set of compatible routes with all these routes being in layout 1.

After having determined the existence of all simultaneities in layout 1 and/or layout 2, the difference in simultaneities between the two layouts is known.

5.2 Check whether layout meets the functional requirements

The tool can check whether a layout meets the functional requirements, specified by the user in terms of reachabilities and simultaneities. It is also possible to require all reachabilities and simultaneities currently available in the layout.

Depending on the length of the list of requirements, the tool will determine all simultaneities and check when finding one if it is in the requirements-list (long list) or it will consider the requirements one by one on the availability of compatible routes (short list).

5.3 Search for redundant infrastructure

Having more points usually means having more functionality, but points are expensive and are a large source of disturbances, so they are assessed more and more critically recently. With the tool the user can search for redundant infrastructure in a layout.

To determine which infrastructure is redundant, the tool needs a specification of the requirements on reachabilities and simultaneities from the user. Besides a list of requirements the user can also require that all current simultaneities in the railway yard must be preserved. With removing the redundant infrastructure, one can come to a railway yard that still has the same functionalities regarding reachabilities and simultaneities, but has less infrastructure. So lower costs and fewer disturbances.

The tool determines which infrastructure is necessary for the requirements. If a simultaneity always needs a certain part of the infrastructure to meet the requirements, that part is necessary. From all the infrastructure that has been



found unnecessary, any link can be removed. However, by removing a link, other simultaneities can now require parts of the infrastructure that were not necessary before, because it could use the removed link before. Therefore, after each removal of a link, the necessary infrastructure has to be recalculated. In this way the user can easily find a smaller layout that still meets the requirements.

Figure 4 shows a railway yard. Figure 5 shows the same railway yard, but with all redundant infrastructure removed. The simultaneities in both layouts are exactly the same (all current simultaneities were required by the user).

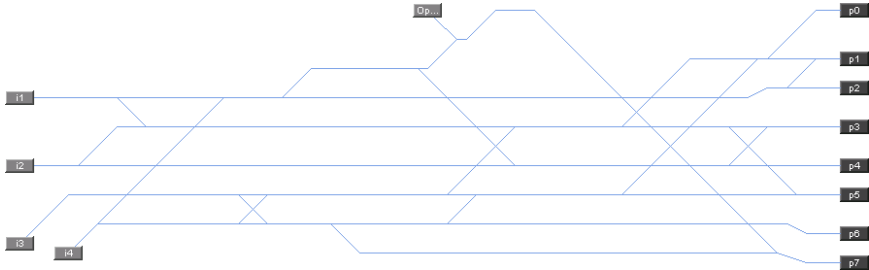


Figure 4: Example of a railway yard.

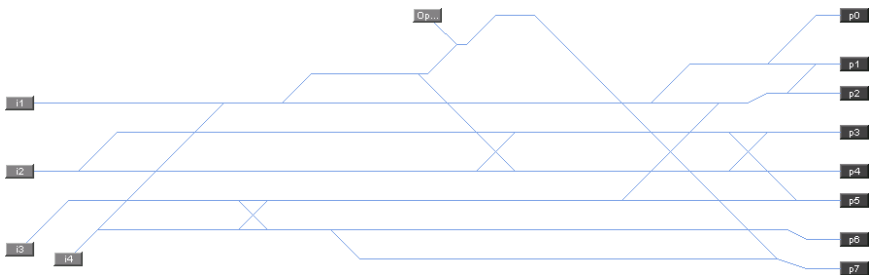


Figure 5: Railway yard of figure 4 without redundant infrastructure.

6 Conclusion

With the tool the possibilities to determine the benefits and necessity of a point have been improved. The tool enables the user to quickly determine which points or links are not necessary to meet the functional requirements and to determine the added functionality when a certain point or link is added.

This means current railway yards or new designs for railway yards can be tested by COMPAT for redundant infrastructure. Until now the results could only be achieved manually, at a much higher costs, longer lead times and lower quality. Therefore, COMPAT provides the railway infrastructure manager with a powerful tool to determine a cost effective layout of a railway yard. By presenting the results graphically, COMPAT contributes to the user's insight into the effects of adding or removing points and links in railway yards, thus reducing discussions among specialists.



References

- [1] Wolbers, M.J., *DSS Gelijktijdigheden*, January 2005 (*in Dutch*).
- [2] Wolbers, M.J., *Decompliceren wisselcomplexen*, June 2005 (*in Dutch*).
- [3] Schrijver, A., *A Course in Combinatorial Optimization*, pp. 149-174, 2004.
- [4] Fortune, S., Hopcroft, J. & Wyllie, J., *The directed subgraph homeomorphism problem*, Theoretical Computer Science 10, pp. 111-121, 1980.



A cost effective solution to manage rail cargo fleets: the final assessment of the F-MAN project

G. Cosulich¹, A. Derito¹, M. Giannettoni² & S. Savio²

¹*Sciro S.p.A., Genoa, Italy*

²*Department of Electrical Engineering, University of Genoa, Italy*

Abstract

The objectives of the EU project Rail Car Asset Management (hereinafter referred to with its short name, F-MAN) were to investigate the feasibility, the potential impacts and benefits of a new approach toward the management of a freight railcars fleet, aiming at improving the competitiveness of Railway Cargo Operators in the context of freight transport. To reach its objectives, the F-MAN Consortium has identified the requirements, designed the architecture, specified and developed the components (on-board and ground fixed) of an IT-based system enabling enhanced fleet management of freight railcars and introducing asset management in the process of the international transport chain. In the paper the authors show the quantitative results of the final assessment of the project, describing also the characteristics of the simulation tool to this aim developed.

Keywords: railway, cargo fleet management, IT-based system, simulation.

1 Introduction

The F-MAN project started in October 2001 from a consortium bringing together the experience, expertise, know-how and knowledge of an heterogeneous set of partners including Railway Cargo Operators (RCOs), developers of Transport and Logistics applications, Information & Communication Technologies experts and Research Institutions [1, 2]. The project was co-funded by the European Community under the Information Society Technologies (IST) Programme and provided its final results in September 2004.

The F-MAN system is made of three main modules, namely the Tracking System Module (TSM), the Data Processing Module (DPM) and the Asset Management Module (AMM). The TSM is constituted by On-Board Terminals



(OBTs), to be installed on each wagon of the managed fleet, and a Ground Station (GS). The On-Board Terminals collect data on wagons position and status by means of GPS receivers and on board sensors, and transmit such data to the GS by means of GSM terminals. Communication is SMS based and event oriented in order to comply with the energy consumption constraints deriving from the maintenance free, long life battery providing the power source.

The GS, which is responsible for OBTs originated/terminated data buffering, pre-processing, conversion and distribution, is connected to the DPM through a secure Internet connection. The DPM is the software application responsible to maintain data relevant to the fleet composition (F-MAN Pool), past, actual and forecasted geographical distribution (Tracking and Tracing and Expected Time of Arrival – ETA), maintenance status (technical availability), and to manage the event messages communication procedures toward the OBTs. Orders processing, wagons selection and booking, trips organisation and logging of data for asset management purposes are performed through the AMM. The AMM and the DPM share a common database and are accessible through a Web Application Server. Any involved user connects himself to the system through the Internet by means of a simple browser. The F-MAN system provides its users with the so called F-MAN tools, i.e. a set of user friendly interfaces to the system (automatic multi language support is implemented) assisting and driving the user throughout any task he's responsible for. The F-MAN tools were developed and integrated into a prototype, whose functions were successfully tested during real operating conditions by monitoring 50 wagons (suitably equipped) from the French, Portuguese and Slovenian railways [3]. Anyway, given the reduced number of tested railcars, a simulation tool was developed within an evaluation Work package, whose primary task was the assessment of the project results effectiveness. In particular, simulations aimed at evaluating the effectiveness of the algorithms and procedures implemented by the F-MAN application when operating on a significant pool of wagons (e.g. thousands of wagons): the scope of such simulations was to gather data enabling a fair comparison between the performances (such as fleet productivity) obtainable from a fleet operated according to the F-MAN approach and those relevant to the nowadays management practice.

2 Simulator characteristics

The simulator was designed and developed to run on the Microsoft Access database engine (Microsoft® Access 2000). It was not intended to emulate the traffic related to wagon movements (there is no representation of railway network layout, capacity, etc.) but only the fleet management approach toward orders fulfilment. To explain such concept with a few words:

- when an order is processed, the engine queries the database for suitable wagons (available in the origin station on departure date), according to predefined criteria;
- if all the wagons have been found, the order is satisfied and wagons will become available again in the destination station at the arrival date.



The following assumptions and requirements have been considered to design and develop the simulation software:

- during the simulation, the total number of wagons in the model is fixed;
- each wagon is identified by a unique wagon identifier and by a wagon owner identifier;
- each station is defined with an initial number of available wagons;
- each station-to-station path is characterised by a mileage;
- trip time for each path is calculated considering an average speed, which may be different for international and domestic trips;
- wagons unavailability time (T_{wu}) can be set as a parameter (e.g., to take into account wagons loading and unloading, or other marshalling operations), so that a wagon can be used for a given task only if the wagon is available for the task at least T_{wu} hours earlier;
- the fleet manager is aware of new orders from any origin to any final destination (inside or outside the origin Country) at departure time $D_0 - \tau_b$, being τ_b a random variable representing the booking time;
- trip times are always the scheduled ones.

Orders lists are generated off-line by Excel[®] macros and then exported to feed the simulator database; any order list is generated for a given number of days, starting from a predefined date. For each origin-destination combination, the orders are generated on the basis of the following criteria:

- a stochastic distribution of wagons per order, from each origin to each destination, is defined to generate the order lists;
- number of orders/day (e.g. 2 if the relevant transport is operated twice a day);
- mean value and standard deviation of the distribution used to extract the number of wagons requested by a single order.

Once lists have been generated, orders are processed by the simulator according to the following criteria:

- orders are processed according to order date;
- orders for the nearest departure date are processed first;
- only orders which can be satisfied 100% are processed;
- each day wagons are booked based on wagons availability and cost of availability;
- for international trips, wagons belonging to the destination Country are selected first;
- re-processing of orders is not allowed.

As far as management models are concerned, it is worth noting that the traditional management model implemented for the simulations takes into account the RIV regime as a reference. Although the EUROP regime would be more similar to the F-MAN pool concept, the Consortium has preferred to carry out the comparison with the RIV regime in order to have the feeling of the potential impacts of the application with respect to the current management policies, being the EUROP regime abandoned.

At last, performance indicators addressed by simulations are listed below (to estimate indicators, the “fleet” is always considered as the whole set of wagons):



- Performance ID 1: fleet productivity computed on a time basis [revenue time/(revenue time + empty time)];
- Performance ID 2: fleet productivity computed on a mileage basis [loaded mileage/(loaded mileage + empty mileage)];
- Performance ID 3: rate of satisfied wagon requests [delivered wagons/requested wagons];
- Performance ID 4: rate of satisfied orders [fulfilled orders/number of orders].

3 Case studies

The authors analysed many different scenarios related to the management of a cargo traffic flow among three areas (Countries), each characterised by four departure/arrival stations.

The first differentiation of simulation scenarios was made by varying the composition of the orders lists to be processed by the simulator. In particular, the following families of orders lists were utilised:

- National 120 days: orders list for 120 days, with approximately 72 orders/day (about 8600 orders and 150000 wagons request), with 75% national trips and 25% international trips;
- International 120 days: roughly the same number of orders and wagons request as above, but with 25% national trips and 75% international trips;
- International 240 days: the same composition as International 120 days, but distributed on 240 days, thus with approximately 36 orders/day.

Further major elements determining the simulation scenario were the wagons fleet size and the distribution of wagons among stations at the beginning of the simulation.

As far as the initial wagons distribution is concerned, test simulations were performed in order to verify that the impact of uneven distributions is limited in time from the beginning of the simulation: this is compliant with the fact that both the Traditional and the F-MAN management approaches have the capability to move wagons where they are needed. However, in order to avoid biasing of simulation results utilised for evaluation purposes, wagons were always equally distributed among stations.

The following fleet sizes were considered in the simulations:

- Large fleet: 4800 wagons equally distributed among the stations (400 wagons/station);
- Reduced fleet (F-MAN only): 3600 wagons (300 wagons/station);
- Small fleet: 2400 wagons equally distributed among the stations (200 wagons/station).

The authors set the average speed for domestic and international trips respectively equal to 30 km/h and 18 km/h, while a value ranging from 4 to 10 days was used for the “send-back” time of RIV wagons (a wagon outside its proprietor’s Country must be repatriated within this time, regardless it can be sent back loaded or not).



As far as the wagons unavailability time (T_{wu}) is concerned, such parameter was set equal to 0 h in all simulations except for those developed to estimate the impact of the accuracy of ETA information on the system performances.

4 Simulation results

In this chapter the authors present some of the most interesting results of the simulations. A summary of the results obtained from the simulations relevant to heavy traffic (National – International 120 days) and undersized fleet (2400 wagons) scenarios is presented in the following Table 1.

Table 1: Summary of the simulation results (heavy traffic – undersized fleet).

	National 120 days – 2400 wagons			International 120 days – 2400 wagons		
	Traditional	F-MAN	$\Delta\%$	Traditional	F-MAN	$\Delta\%$
Performance ID 1	0.77	0.95	+ 24%	0.84	0.98	+ 16%
Performance ID 2	0.74	0.94	+ 28%	0.79	0.97	+ 23%
Performance ID 3	0.72	0.96	+ 34%	0.54	0.75	+ 38%
Performance ID 4	0.74	0.97	+ 31%	0.58	0.80	+ 37%
Loaded mileage [Mkm]	51.8	70.9	+ 37%	57.6	79.6	+ 38%
Empty mileage [Mkm]	18.3	4.4	- 76%	15.7	2.7	- 83%
Wagons cycle time [h] ([days])	66.7 (2.8)	50.0 (2.1)	- 25%	88.4 (3.7)	64.3 (2.7)	- 27%

Such table shows, for the Traditional management approach and the F-MAN pool management approach, the following data, calculated as mean values over 5 replications for each orders list composition:

- performance indicators, as defined in the previous Section 2;
- loaded mileage of the fleet (sum of all the km relevant to movements of loaded wagons);
- empty mileage of the fleet (sum of all the km relevant to movements of empty wagons);
- fleet wagons cycle time (T_{wc}), considered as the mean time between two successive missions of a wagon and computed as it follows:

$$T_{wc} = \frac{N_w T_s}{\sum_{i=1}^{N_{so}} N_{wi}} \tag{1}$$

where:

N_w = total number of wagons

T_s = total simulation time



N_{so} = number of satisfied orders

N_{wi} = number of wagons for the i -th satisfied order

Difference (in percentage) between the F-MAN values and the Traditional (i.e., reference case) values are reported as well.

It is straightforward noting that the F-MAN pool management approach always achieves better performances than the Traditional management approach:

- F-MAN enhances the exploitation of the fleet, as the loaded mileage (payload) increases;
- F-MAN improves the productivity of the fleet, as the increase of the loaded mileage is achieved with a striking reduction of the empty mileage.

Although the F-MAN approach had been conceived to improve cargo fleet performances for international traffic flows, it is worth noting that the benefits are remarkable also in scenarios characterised by a majority of domestic traffic (National). In fact, despite international traffic weight is just around 25% of the total traffic in the National orders list composition, the impact of the F-MAN approach is estimated as about a 30% increase in the capability of satisfying orders requests, contemporaneously reducing the movements of empty wagons to about 25% of the empty movements needed in the Traditional approach. This is a result that goes over and above the most optimistic expectations.

Simulation results show that, when the traffic scenario is shifted from National to International, performance indicators relevant to loaded time and mileage increase, while the empty mileage and the indicators relevant to the fulfilment of orders request decrease, for both management approaches.

Such behaviour is a direct consequence of the characteristics of the International traffic scenario analysed and of the fleet size (undersized). In particular, moving from the National orders list to the International one, the frequency of orders request for transport from a given station to stations of other Countries significantly increases, with an increase of both the mean trip distance and the mean time to accomplish an order.

Moreover, the send back time (RIV regime) for those simulations was set to 10 days, a period during which each station has plenty of orders to be fulfilled (most of them to foreign Countries for the International orders list). Under those conditions, each Country has many occasions to load and send a foreign wagon to the belonging Country, without any need for sending it empty just to comply with the send back time.

All the considerations mentioned above clearly testify the reasons why the benefits deriving from the adoption of the F-MAN approach in the case of the International scenario, although definitely huge, are analogous to the ones related to the National scenario, as the operating conditions are in the latter case favourable on the side of the Traditional management.

The second set of results presented by the authors deals with the simulations carried out in order to estimate the impact of ETA accuracy on the performances of the F-MAN Pool management approach as implemented by the simulator model. The estimate has been carried out by means of a sensitivity analysis on the T_{wu} parameter, i.e. the value of the wagons unavailability time. As mentioned



before, this is the time during which a wagon, following the completion of its mission, is not yet available. Such parameter, originally introduced to take into account the loading/unloading time, has been used to take into account the ETA accuracy instead. Despite representing a rough – but working – method expected to penalize the performance more than in the real implementation of the system (the relevant unavailability time is applied after any completed mission), this allows to evaluate what the performance would be (in the worst case) when the F-MAN AMM considers a time equal to T_{wu} as the accuracy of the ETA information. Results summarised in Table 2 for simulation scenarios characterised by heavy traffic and sufficient fleet (to fulfil all the orders when T_{wu} is set to 0) show that when the accuracy of ETA decreases (simulated through an increase of T_{wu}), the performances decrease as well, since the number of rejected orders and undelivered wagons increase, while the loaded mileage decreases. It is worth noting that the slight increase of performance indicators ID 1 and ID 2 is only due to the lower number of running wagons as a consequence of the lower number of satisfied orders.

Table 2: Impact of T_{wu} parameter (heavy traffic – sufficient fleet).

	International 120 days – 3600 wagons			
T_{wu} [h]	0	6	10	18
Performance ID 1	0.965	0.966	0.969	0.973
Performance ID 2	0.958	0.958	0.960	0.962
Performance ID 3	1	0.986	0.967	0.894
Performance ID 4	1	0.994	0.981	0.925
Loaded mileage [Mkm]	107.1	106.4	104.5	96.3
Empty mileage [Mkm]	4.6	4.7	4.3	3.8
Wagons cycle time [h] ([days])	71.2 (3)	71.7 (3)	73.1 (3)	79.1 (3.3)
Rejected orders	0	41	139	547
Undelivered wagons	0	865	3129	12118

In the last section of this chapter the authors present the results of the analysis carried out to evaluate the potential impact of the F-MAN Pool management approach on the productivity of a fleet of wagons, taking into account economical issues as well. It is worth noting that the indicators previously utilised to score the performance of the management approaches are not sufficient to this aim; thus another indicator linking together those already mentioned with the costs of owning and operating the fleets is introduced.

The indicator utilized by the authors is the daily wagon productivity (P_{wd}), defined as follows, respectively for the Traditional and the F-MAN wagon (subscript -T and -F applies for Traditional and F-MAN fleet respectively):

$$P_{wd-T} = [R_{lt} N_{kml-T} - (C_{wd} N_{w-T} T_D + C_{et} N_{kme-T})] / (N_{w-T} T_D) \tag{2}$$



$$P_{wd-F} = [R_{lt} N_{kml-F} - (K_F C_{wd} N_{w-F} T_D + C_{et} N_{kme-F})] / (N_{w-F} T_D) \quad (3)$$

where:

N_w = total number of wagons in the fleet

N_{kml} = total loaded mileage of the fleet [km]

N_{kme} = total empty mileage of the fleet [km]

T_D = observed time period [days]

C_{wd} = daily cost of owning a wagon (used or not) [€/day]

C_{et} = cost per km of moving an empty wagon [€/km]

R_{lt} = net revenue per km for a loaded wagon [€/km]

K_F = F-MAN equipment factor (>1, takes into account the daily cost of OBT and communication)

Being the values of C_{wd} , C_{et} and R_{lt} , in the real context, highly variable from RCO to RCO and from wagon type to wagon type, the analysis was carried out estimating the values of the daily wagon productivity in per units of the daily cost of a wagon (C_{wd}), defined as:

$$p_{wd-T} = [C_1 N_{kml-T} - (N_{w-T} T_D + C_2 N_{kme-T})] / (N_{w-T} T_D) \quad (4)$$

$$p_{wd-F} = [C_1 N_{kml-F} - (K_F N_{w-F} T_D + C_2 N_{kme-F})] / (N_{w-F} T_D) \quad (5)$$

where:

$C_1 = R_{lt} / C_{wd}$

$C_2 = C_{et} / C_{wd}$

To carry out the analysis K_F was set to 1.1, considering a daily wagon cost of about €15 (a mean value derived from the information provided by users within the Consortium) and the following data related to F-MAN equipment:

- OBT provision and installation equal to about €2500;
- OBT life (maintenance free) equal to 5 years;
- mean daily communication cost per wagon equal to about €0.12.

In Figures 1 and 2 the authors present the results related to the data collected during the simulation of the heavy (International 120 days) traffic – sufficient fleet scenario (send back time for the RIV regime set to 4 days).

It is worth noting that, due to the different performances of the F-MAN management with respect to the Traditional one, the need for a suitable number of wagons able to satisfy almost all the orders resulted in an F-MAN fleet size smaller than the Traditional one (3600 and 4800 wagons respectively). In particular, Figure 1 shows the per cent daily wagon productivity increase of F-MAN with respect to the Traditional management as a function of C_1 and C_2 , in the ranges [0.06, 0.1] and [0.02, 0.04] respectively. Figure 2 presents three cross sections of the previous Figure 1, i.e. the per cent daily wagon productivity increase of F-MAN with respect to the Traditional management as a function of C_1 for three fixed values of C_2 .



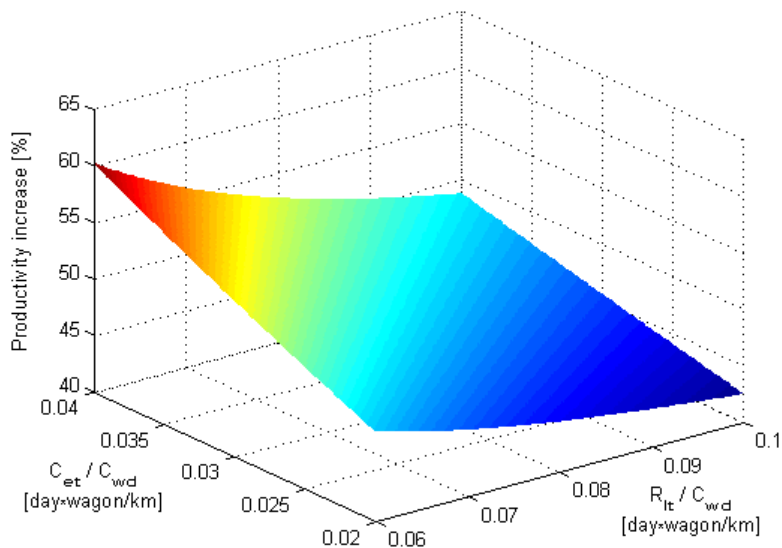


Figure 1: Daily wagon productivity increase with F-MAN.

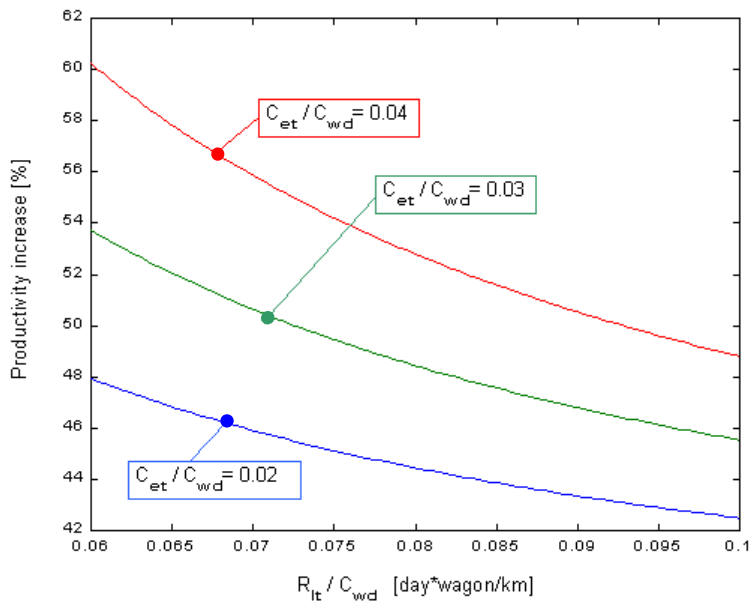


Figure 2: Daily wagon productivity increase with F-MAN (fixed C_{et}/C_{wd}).

The reported results clearly show that the F-MAN Pool management approach is estimated to significantly improve the productivity of the fleet. Depending on fleet specific costs, the percentage increase of the daily wagon productivity may range from approximately 40% up to 60% and more. Although this is quite a wide spread, such result is definitely a very positive one.



Of course the actual costs borne by the RCO to implement and utilise the F-MAN system will largely depend on commercial agreements with OBT suppliers and mobile communication providers, so that the actual value of K_F for a specific wagon type may differ from 1.1.

For such reasons, the authors carried out a sensitivity analysis in order to evaluate the impact of an increase of the parameter K_F on the effectiveness of the F-MAN solution. Such final analysis demonstrated that the impact of the costs of the F-MAN system on fleet productivity is very limited.

To give an example, in the relaxed (International 240 days) traffic – oversized fleet scenario (3600 and 4800 wagons for the F-MAN fleet and the Traditional one respectively), which resulted the worst case as far as fleet productivity is concerned, even if K_F equals 2 (i.e. the daily cost per wagon of the overall F-MAN system equals the daily cost of owning a wagon) the productivity increase would still be around 45% (send back time for the RIV regime set to 4 days).

5 Conclusions

The main results of the final assessment of the EU project F-MAN have been presented by the authors in this paper. F-MAN developed and tested a web based IT application providing a set of tools enabling a new approach to fleet management in railway freight transports. The F-MAN pool regime implementation would allow Railway Cargo Operators to gain 2 digits productivity increase on both national and international transports. Comparisons made with other management regimes, such as RIV, show the productivity increase can be estimated between 40% and 60%, or even more.

Again, it has to be reminded that such results have been obtained through simulations implementing simplified models of the management processes, but a qualitative analysis of the results allows to conclude that the development of the F-MAN system as an industrial quality product, together with the adoption of the F-MAN Pool approach, show concrete potentialities to represent a very good deal for both developers and users.

References

- [1] Bozzo, R., Derito, A., Kuhla, E., Nurchi, R. & Recagno, V., A vision of future freight rail fleet management, *Proc. of the World Congress on Railway Research WCRR 2001*, Köln, Germany, 25-29 November 2001.
- [2] Ballis, A., Guigon, M. & Kuhla, E., F-MAN will boost wagon productivity in Europe, *Railway International Gazette*, November 2003 issue, pp. 701-703.
- [3] Giannettoni, M. & Savio, S., Fleet management in railway freight transport: the EU project F-MAN, *Proc. of the Int. Conf. COMPRAIL 2004*, Dresden, Germany, 17-19 May 2004.



Section 17

Condition monitoring

This page intentionally left blank

Methodology for the monitoring, control and warning of defects for preventive maintenance of rails

K. C. Labropoulos, P. Moundoulas & A. Moropoulou

*Materials Science and Engineering Laboratory,
School of Chemical Engineering,
National Technical University of Athens, Greece*

Abstract

The preventive maintenance of rails is necessary for safe, successful and economical railway performance and operations. Rolling contact fatigue (RCF) is one of the main issues that concern the head of the railway. Progressively, RCF defects can propagate inside the material with the risk of damaging the rail, and thus they need to be monitored, categorized, evaluated, and treated through a comprehensive maintenance system. In a previous work with the Athens Metro, an innovative monitoring system was developed based on the capability of non-destructive techniques, such as IR thermography, to inspect and identify defects at early stages and reveal the activated paths of fracture (Moropoulou A., Avdelidis N., et al. *Thermosense XXVII*, 5782, 371-378, 2005). In the present work with the Athens Piraeus Electric Railways S.A. this system is further integrated into a complete system consisting of three components. The surveillance component identifies the rail defects and includes non-destructive techniques, such as visual inspection, infra-red thermography, ultrasonics, fiber optics microscopy, and ACFM. The data management component receives information from the various defect identification techniques, categorizes them and stores them for further use. The defect warning component uses risk indices such as defect level indices, defect extent indices, and risk threshold values, in order to evaluate the significance of the observed defects. All three components are part of a decision making system that monitors the development of defects, warns for significant threats and schedules rail maintenance.

Keywords: preventive rail maintenance, rail defect management, risk indices.



1 Introduction

The preventive maintenance of rails, one of the most critical components of a railway system, is necessary for safe, successful and economical railway performance and operations. The rail distributes the forces from the train to the track bed and together with the tread and flange of the wheel allows for the movement of the train. The forces exerted on a rail include the vertical load and the directional and frictional components as the train moves.

The interaction between the wheels and the steel rail, and in particular the wheel - rail interface, has been the subject of intense research since it controls the initiation and development of defects. In order to be effective, a preventive maintenance system has to take into account the decay mechanisms and be able to reveal the extent of rail damage, so that the appropriate remedy measures can be taken. The overall process is termed "Rail Defect Management" (RDM) and it will be described below.

2 Rail defects

The wear resistance of rails is mainly controlled by hardness. However, the wear resistance is also dependent on the stresses the rails are subjected to that include bending and shear stresses, contact stresses, thermal stresses, and residual stresses. These stresses control the development of defects in rails, eventually leading to failure Cannon et al. [1]. Bending stresses act either vertically (vertical load of the vehicle), or laterally (dynamic and directional forces). The vertical bending stresses are mainly compressive in the rail head and tensile in the rail base. Contact stresses originate from the wheel load, traction, braking and steering actions. Thermal stresses originate from the welding process, during the connection of rail sections to create a continuously welded rail, whereas, residual stresses originate from the manufacturing processes.

2.1 Rolling contact fatigue

Various types of failures in steel rails exist that originate from various internal and intrinsic and extrinsic factors [1]. Such failures are generally categorized as: (a) *Manufacturing defects: they originate from the manufacturing process of the rail;* (b) *Inappropriate handling, installation and use: defects originating from out-of-specifications installation of rails, unexpected scratches and wheelburn defects;* (c) *Decrease of the metal's resistance to fatigue: They include the most common rail defects such as head checks and squats.*

A well designed railway system can minimize the effect of the first two types of defects. The defects due to the decrease of the metal's resistance to fatigue cannot be avoided, and the railway system can only either delay their occurrence or minimize their effect on safe operations. The term "Rolling Contact Fatigue" (RCF) refers to the fatigue of the metal of the rail due to the application of repeated and of high-values vertical and tangential stresses, and describes the initiation and development of rail defects.



When the wheel comes in contact with the rail head, both are deformed and they effectively create two “waves” at the leading and trailing edges of the wheel-rail interface, that travel along the wheel as the later moves on the rail head. Böhmer et al. [2] have shown that these deformed areas as well as the area below the wheel-rail interface are subjected to a complex stress distribution that shears a surface layer, the depth of which depends on the history and remainder properties of the metal.

The shearing of the surface layer of the rail initiates microscopic cracks. The application of repeated loadings has the effect of these microscopic cracks developing further and propagating under small angles through the plastically deformed surface layers [3]. This occurs down to a certain depth ($\sim 5\text{mm}$) that depends on the loading pattern and the material used, until steel exhibits its isotropic properties. Generally, these microcracks develop in a direction parallel to the running direction of the vehicle.

2.2 Classification of defects

Classification of defects is typically performed by use of a catalogue of defects, such as the *UIC (2002) Handbook of rail defects - UIC Code 712R (4th Edn.)*. Many types of RCF related failures exist, of which the two main ones are “head checks” and “squats” [1]. “Head checks” are defined as low angle microcracks with the same direction as the running direction of the vehicle that develop near the gauge corner of the rail and correspond to repeated plastic deformation beyond the plastic shakedown limit. A large number of head checks in a limited area may lead to spalling, or in most extreme cases, propagation of the microcracks in a transverse direction, leading to rail fracture. “Squats” are identified as a broadening of the wheel-rail contact band and occur on the surface of the rail head.

3 Monitoring of rail defects

The detection of rail defects is an active research field and a wide range of non-destructive techniques have been employed with varying success. A simple method of detecting defects in track is the visual inspection carried out by track maintenance staff. Often this is the only method used by some railway operators, due to limited financial resources. However, advanced systems using CCD cameras and laser profilometers, linked to digital video recorders are fielded by modern operators, enhancing the capabilities of the method.

Ultrasonic defect detection has been widely used for the past forty years to reveal surface breakings and internal defects. Unfortunately, despite the advent of wheel array probes with multiple transducers operating at different frequencies and angles, often, small surface breaking defects cannot be reliably detected. Vehicle based ultrasonic test systems operate typically at $40\text{--}70\text{ km h}^{-1}$, although the average speed is often much lower when manual verification is required, further limiting the effectiveness of the method for large rail networks.



Other methods to detect defects that are currently used or under development include eddy current testing, electromagnetic acoustic transducers, radiography, ground penetrating radar, laser generation and reception of ultrasonic waves, alternating current potential drop (ACPD), alternating current field measurement (ACFM), and impedance spectroscopy [1, 4].

In a previous work funded by the Athens Metro, an innovative monitoring system was developed based on the capability of non-destructive techniques such as infrared thermography and fibre optics microscopy to inspect and identify defects at early stages and reveal the activated paths of fracture [5, 6].

4 Preventive maintenance system for Athens Piraeus Electric Railways S.A

The Athens Piraeus Electric Railways S.A. (ISAP), has been operating a Metro line in Athens, Greece for the past 136 years. Over 400,000 passengers use the specific line daily, amounting to over 80 millions annually. The company owns 25.6 km of double rail tracks, using K39, S49, UIC54 and UIC60 rails with wooden and full-body prestressed concrete sleepers type B70. ISAP is currently initiating a significant infrastructure upgrade that will include the complete replacement of older types rails and sleepers. As part of this modernization plan, ISAP has funded a research program with the National Technical University of Athens, School of Chem. Eng., Lab. of Materials Science and Eng., to provide specifications and methodology for the development of system for monitoring, control and warning of defects for the preventive maintenance of rails [4].

4.1 Rail defect management

Rail defects cause significant costs to railway operators. These include [1]: (a) inspection costs - depending on the frequency of inspection; (b) maintenance costs - rail replacements, weld repair, use of lubricants; (c) pre-emptive treatments - rail reprofiling to hinder crack growth; (d) derailments; (e) indirect costs - loss of business confidence and customer support due to safety issues or unnecessary delays. As a result, railway companies try to minimize these costs through regular maintenance of the rails and through appropriate remedial actions when defects are detected.

Rail Defect Management (RDM) is the overall process that aims to identify, classify, assess and remedy defects before they lead to rail failures, and to minimize and if possible avoid rail defects [1]. In order to accomplish these objectives an effective RDM structure consists of two basic stages:

1. Rail inspection which includes (a) detection of defects with non-destructive techniques; (b) classification of defects; (c) decay mapping, to evaluate the extent and rate of development of wear; (d) study of decay mechanisms
2. Treatment of decay problems which includes (a) risk assessment; (b) assessment of intervention necessity; (c) application of limitations; (d) intervention - lubrication / rail reprofiling / rail replacement



4.2 Surveillance system

Detection of defects, especially at their early stages, plays a crucial role in an effective RDM. A wide range of non-destructive techniques, as mentioned above, are typically employed to identify rail defects. These techniques are either used as independent systems, or as subsystems of an integrated surveillance system mounted on a special vehicle or train. The selection to use non-destructive techniques independently or integrated in a complex surveillance system depends heavily on the available human and financial resources, and the needs of each railway company. For ISAP, with the rather limited rail network, a modular surveillance system is specified that consists of up to seven subsystems, each of which can be integrated gradually as the needs arise, as experience is gained, and as financial resources allow.

Specifically, the visual inspection (VI) subsystem can consist of up to one high-speed CCD camera and two laser profilometers per rail, the digital signal of which will be sent to a digital video recorder and an LCD. The infrared thermography (IRT) subsystem will consist of a 8-14 μm IR camera, with at least 320x256 pixels resolution and better than 0.08°C thermal sensitivity. The IRT subsystem should be supplemented by a fully developed thermal imaging and analysis software, a digital video recorder, portable PC, and a thermal excitation source for use during the night.

A fiber optics microscopy (FOM) subsystem can be used in conjunction with the IRT and VI subsystems, to aid in the in-situ identification and the acquisition of high resolution micrographs of defects for further analysis. It can consist of a walking stick configuration employing a CCD camera with at least 1024x768 pixels resolution and zoom lenses 25x-200x with autoscale feature.

Due to its importance and extensive experience in railway applications, an ultrasonic testing (US) subsystem is included that will be portable (optionally vehicle-mounted if desired). The US subsystem will consist of a UT wheel array probe operating at 2-20 MHz, offering A-scan, B-scan and C-scan capabilities. Similarly, an ACFM subsystem can be included based on international experience with the technique. It will also have a walking stick configuration with an option for vehicle mount, and will consist of an array of 8 pairs of Bx/Bz coils, at 50 kHz. Both subsystems can be linked to the same operator's station.

The above surveillance subsystems will be supplemented by a number of support subsystems. Specifically, a precision positioning / navigation (PPN) subsystem is considered necessary for decay mapping, since its capabilities control the resolution with which the development of defects with time can be monitored. Advanced PPN subsystems can include inertial navigation systems (INS), linked to DGPS, event triggers, and to the main train position control unit. Preferred position accuracy is less than 1m. A rail cleaning subsystem consisting of brushes, heating units and blowers will also permit all-weather operations.

All subsystems briefly described above can be either used independently (portable) or preferable mounted on a special vehicle. This can either be a truck able to travel on rails or an appropriately modified train. ISAP will initially field a selected number of surveillance subsystems independently and once experience



is gained the company will upgrade to more expensive options. An ISO container with three operator's stations, mounted on a truck with detachable mounting for an array of surveillance subsystems has been proposed to ISAP for future use.

4.3 Data management system

As described above, the defect detection techniques can be used independently, and this is typically the case with the majority of railway operators that do not have the ability to procure comprehensive integrated rail defect detection vehicles (e.g. specialized trains). For most Metro lines, such as ISAP, to resort to such expensive solutions is not a viable option. In fact, a great improvement in the effectiveness of the available surveillance subsystems can be achieved by their gradual or partial integration into a data management system, which with an open architecture can be developed to any required level of complexity.

The basic features of such a data management system should include:

- Ethernet network
- Integration of surveillance subsystems into a local network with standard analog, digital or serial input/output units
- Development of a specialized database
- Computer/server which can be configured either (a) for in-situ, real time data acquisition / data storing, and in-situ or remote, near real time data processing; or (b) for in-situ, real time data acquisition / data storing, and post-processing of data at the technical base. The use of GSM Railway mobile communication networks can be a valuable tool.

4.4 Defect warning system

Non-destructive detection of rail defects and storage of the acquired data are important capabilities for a maintenance system of a railway company, but if they are not combined with an expert rail defect warning system (ERDWS) their effectiveness is limited. The lack of an ERDWS essentially restricts the preventive maintenance system only to knowledge of the extent of rail decay. Any remedy actions are made through traditional decision making procedures involving past experience of the technical staff and available resources. As a result, rails are often replaced either too early or too late. In contrast, an ERDWS supports all stages of preventive maintenance after the rail decay has been identified, which are described below.

4.4.1 Archiving and classification of rail defects

Once detected, classification of rail defects should follow the general guidelines of standard defect classification catalogues such as the *UIC (2002) Handbook of rail defects - UIC Code 712R (4th Edn.)*. Archiving of each defect in a database should include further information, such as a unique identification number of the defect, type of defect (4-digit number), date of detection, exact location relative to a reference system, detection technique, link to relevant images/micrographs/thermographs/data, environmental data (temperature of the environment, temperature of rail, relative humidity), rail identification number,



and any other notes made by the maintenance operators or the train drivers. The database should be sufficient large to store all available information and should be accessible by the in-situ operators (e.g. through GSM-R)

4.4.2 Defect level indices

The term “defect level index” refers to the evaluation of the state of a rail section or part thereof and corresponds to the worst (most dangerous) value of every observed type of defect. Specifically, appropriate software will search through the database to reveal all observed types present on a specific rail section or part thereof, and for a specific date. Then, the observed defects will be grouped in defect categories (e.g. according to UIC code 712R) and relevant defect level indices will be calculated. Table 1 shows a selection of defect level indices that can be defined to describe defects observed on the head of a rail. The numerical index i is the unique identification of each observed defect.

Table 1: Typical defect level indices for defects on the rail head.

Index	Description
$L_i^{trans-head}$	Length of transverse crack of the head [111, 211, 100, 200]
$L_i^{horiz-head}$	Length of horizontal crack of the head [112, 212]
$L_i^{long-head}$	Length of longitudinal, vertical crack of the head [113, 213]
$L_i^{corrugation}$	Length of corrugations [2201, 2202]
$\lambda_i^{corrugation}$	Wavelength of corrugations [2201, 2202]
$D_i^{trans-head}$	Depth of transverse crack of the head [111, 211, 100, 200]
$W_i^{horiz-head}$	Width of horizontal crack of the head [112, 212]
$D_i^{long-head}$	Depth of longitudinal, vertical crack of the head [113, 213]
$A_i^{surf-defect}$	Area of surface defect [121, 221]
$A_i^{shell-run}$	Area of shelling of running surface [2221, 122]
$A_i^{shell-gauge}$	Area of gauge corner shelling [2222]
A_i^{crush}	Area of crushing [123, 223]
A_i^{burn}	Area of wheel burns [125, 2251, 2252]
A_i^{uneven}	Area of local unevenness [124, 224]
$A_i^{corrugation}$	Area of corrugations [2201, 2202]



4.4.3 Defect extent indices

Although the defect level indices reveal the most dangerous situations for each type of defects, it is often desirable to know the extent of decay of the rail. The term “defect extent index” is used to describe the percentage of occurrence of each type of defect relative to a rail section or part thereof. If the reference system is a rail section, then the defect extent indices reveal the total percentage of occurrence of each type of defect for a given rail section length, l_{rail} . If the reference system is a specific part of a rail, with a length l_x , then the defect extent indices correspond to the “concentration” of each type of defect in the given rail part, usually of different value than those for the whole rail section. Table 2 shows a selection of typical defect extent indices defined for a rail section, l_{rail} . Similar defect extent indices can be defined for other rail defect types.

Table 2: Typical defect extent indices for defects on the rail head.

Index	Description
$L_{spec-rail}^{trans-head}$	Specific length of transverse cracks of the head [111, 211, 100, 200] $= \sum_{i=1}^n \left(\frac{length\ trans\ crack\ i}{rail\ width} \right)$ where n is the total number of transverse cracks of the head for the given rail length l_{rail}
$P_{mean-rail}^{trans-head}$	Mean percentage of transverse crack of the head [111,211,100,200] $= \sum_{i=1}^n (area\ trans\ crack\ i) \cdot 100 / (area\ rail\ cross - section)$ where n is the total number of transverse cracks of the head for the given rail length l_{rail}
$P_{mean-rail}^{shell-run}$	Mean percentage of shelling of running surface) [2221, 122] $= \frac{\sum_{i=1}^n (area\ shelling\ i\ at\ running\ surface) \cdot 100}{area\ rail\ head}$ where n is the total number of shelling defects at the running surface for the given rail length l_{rail}

4.4.4 Risk assessment

Assessment of the risk of rail failure for the observed defects will be done automatically by a software that will use the defect level indices and defect extent indices and compare them to corresponding values from a risk threshold table. The latter will contain the threshold values for each index, which will be defined experimentally by “calibration” curves that correlate the probability of failure as a function of the observed defect index.

As an example, Figure 1 shows the relevance of the threshold values for the defect level index $D_i^{trans-head}$, the depth of a transverse crack i at the head of a rail.



The left curve corresponds to the tensile strength of a rail as a function of the depth of a transverse crack on the rail head. The right curve corresponds to the probability of failure of the rail as a function of the depth of a transverse crack i at the head of the rail. If two probabilities are selected to define the desired ranges of risk of failure (see horizontal dashed lines), Three “regions of risk state” can be calculated by the interception of the respective horizontal lines with the probability of failure curve, thus providing the corresponding threshold values. Similarly, threshold values can be defined for all other indices.

Once the threshold values are defined, the defect level indices for each classified defect and the defect extent indices for a specific rail section or part thereof will be compared to the threshold values and the worst case will correspond to the risk state of each individual rail section or part thereof.

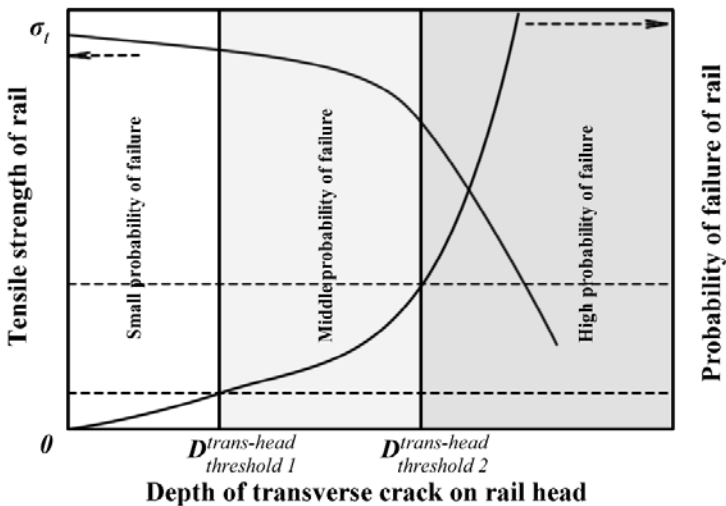


Figure 1: Schematic representation of the calculation of threshold values for the defect level index $D_i^{trans-head}$ and the definition of risk regions.

4.4.5 Scheduling of maintenance

The next stage, after the risk assessment, is to schedule rail maintenance. Often, although an intervention is justified due to the risk of failure involved, the final scheduling of maintenance has to take into account other factors such as, safety issues, intervention costs, cost of alternative options, remaining life of rail, schedule of train routes, existing know-how, availability of maintenance tools and materials, etc. In some cases, especially at low probabilities of failure, limitations to traffic and/or loads are imposed, until more permanent measures are taken to treat the problem. Thus, the ERDWS should be able to provide effective guidelines for scheduling of a preventive maintenance and for corrective actions. The final decisions can be made by the technical staff.



5 Conclusions

In the present work, the methodology for the development of a modular system for the monitoring, control and warning of defects for preventive maintenance of rails has been presented, that includes a surveillance component, a data management component and a defect warning component that uses risk indices. All three components are part of a decision making system that monitors the development of defects, warns for significant threats and schedules rail maintenance.

Acknowledgement

The authors would like to thank Athens Piraeus Electric Railways S.A. for funding the present work.

References

- [1] Cannon, D.F., Edel, K.-O., Grassie, S.L. & Sawley, K., Rail defects: an overview. *Fatigue & Fracture of Engineering Materials & Structures*, **26**, pp. 865-886, 2003.
- [2] Böhmer, A., Ertz, A. & Knothe, K., Shakedown limit of rail surfaces including material hardening and thermal stresses. *Fatigue & Fracture of Engineering Materials & Structures*, **26**, pp. 985-998, 2003.
- [3] Ringsberg, J.W., Life prediction of rolling contact fatigue crack initiation. *International Journal of Fatigue*, **23**, pp. 575-586, 2001.
- [4] Moropoulou, A. *et al.* *Specifications and methodology for the development of system for monitoring, control and warning of defects for the preventive maintenance of rails*, Technical report to Athens Piraeus Electric Railways S.A, National Technical University of Athens, 2006
- [5] Moropoulou, A. *et al.* *Development, application, standardization and installation of a non-destructive testing system for the preventive maintenance and optimization of rail infrastructure life*, Technical report to Attiko Metro S.A., National Technical University of Athens, 2004
- [6] Moropoulou, A., Avdelidis, N.P., Labropoulos, K., Moundoulas, P., Polymenakou, D., Stavarakas, D. & Terezakis, P. Use of thermography and fibre optics microscopy for the evaluation of railway structures in the Athens Metro in Greece, *Thermosense XXVII, Vol. 5782*, eds. G.R. Peacock, D.D. Burleigh, J. Miles, Vol. 5782, SPIE Publ., pp. 371-378, 2005



Onboard measurement method for signaling equipment on Probe Trains

H. Nakamura¹, S. Takahashi¹, T. Hiramoto¹, H. Mochizuki¹
& T. Mizuma²

¹*Department of Electronics and Computer Science,
Nihon University, Japan*

²*National Traffic Safety and Environment Laboratory, Japan*

Abstract

An automatic train protection system ATS and a track circuit is a fundamental and essential element of a railway signaling system. At first, we mention maintenance technology of ATSS. For the stable operation of AN ATS, the management of the quality factor (Q) of the ground beacon is essential and the Q must be kept its value within A specified level through maintenance. Our developed new measurement method is based on the current ratio between two different frequency levels of a pick-up coil current when the train passes the beacon. And it can provide the correct Q data against the distance-fluctuation. Secondly we develop the onboard method of measuring track circuit. The current of track circuit which is received through a pair of onboard signal receivers is continuously recorded and analyzed. From the analyzed data signal engineers judge the necessity of maintenance of track circuits. These measuring technologies are installed onto the probe train which is one of revenue-earning train with measurement functions. A condition-based monitoring system is realized through the result of statistic processing of monitored data which is collected from two or more probe trains.

Keywords: signaling system, ATS, track circuit, quality factor, probe train.

1 Introduction

Reduction of maintenance cost is an important task for railroad transportation. The railroad operators are making efforts to reduce their maintenance costs through modernizing and laborsaving of maintenance. Since measurement by instruments installed along the railroad and measurement of track conditions require a lot of



manpower, automatic measurement by an inspection car sometimes takes place. However, inspection cars are expensive, and because they must be run on an extra schedule, there are restrictions on their running. The inspection cars owned by the railroad operator are not sufficient. To ameliorate such a situation, we are developing a probe train [1–3]. The probe train is a commercial train that uses simple sensors to automatically measure the conditions of the track and signal facilities making up the railroad system. This article introduces the measurement functions offered by the probe train, or the measurement technology of ATS (automatic train stopping system) and track circuits, among others.

2 Probe technology for ATS

2.1 Problems faced by ATS maintenance and background of technical development

If the railroad crew should overlook a stop signal indication and miss brake handling, a safety system automatically operates the brake to protect the safety of the train. Automatic train stopping systems (ATS) are examples of that kind of system. The ATS introduced on all JR lines and many other lines in Japan is a frequency shift ATS system. The frequency shift ATS uses an LC coil called a beacon installed between the rails. When a train passes over the coil, the ATS receives the signal indication from the ground. Thus it is a point control ATS. If reception of information from the point control ATS becomes impossible, it is very dangerous, because it leads to a loss of function. Therefore, quality control of beacons has become an important maintenance item. For the beacons, their quality control has been handled by the Q -value (quality factor) of the LCR series circuit, and their performance measurement has taken place with a fixed period on the field or from the train. The Q -value of ATS beacons placed sporadically along the railroad has to be measured one by one using a Q -meter. This method of measurement is exact but requires much time and manpower, so it has been considered one of the labor-saving problems of signal maintenance.

On the other hand, automatic Q -value measurement also takes place using an inspection car. Since the signal reception level when a train passes over the beacon is proportional to the Q -value, the Q -value can be determined by measuring the reception level by an inspection car. However, because the reception level depends not only on the Q -value, but also on the relative distance from the beacon to the on-board pick-up coil and the installed condition of the beacon, measurement error has been a problem.

The authors are developing probe train technology aiming at monitoring the railroad facility condition by a commercial train. For measurement of the Q -value of the ATS beacon by this probe train, it is indispensable to establish a measuring technique to measure the Q -value in the shortest possible time and with high accuracy. Section 2 of this article reports the results of our examination and experiments.



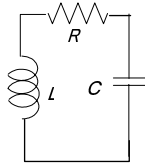


Figure 1: Equivalent circuit of ATS beacon.

2.2 Quality control of ATS beacon

The ATS beacon transfers information by pulling the on-board transmission frequency to the resonant frequency of the beacon. There are about 7 frequencies available nowadays, and information is obtained by discriminating the shifted frequencies through a filter on board. For this frequency shift to take place for certain, the Q -value of the beacon must be larger than a certain value. In maintenance work, therefore, for quality control of the beacon, the Q -value measurement of the beacon is performed in addition to appearance inspection by the eye.

2.3 ATS beacon and Q -value

The ATS beacon can be expressed as the LCR series circuit shown in Figure 1. The angular resonant frequency ω_0 of the LCR series circuit can be obtained from Equation (1) by setting the imaginary part of the impedance to zero.

$$\omega_0^2 = \frac{1}{LC} \quad (1)$$

From the condition that the absolute values of the real part R and the imaginary part $(\omega L - (\omega C)^{-1})$ are equal, the condition of Equation (2) is obtained. The cut-off angular frequencies $\omega_b(\omega_{b1}, \omega_{b2})$ are obtained as the solutions of Equation (3).

$$R = \left| \omega_b L - \frac{1}{\omega_b C} \right| \quad (2)$$

$$LC\omega_b^2 + RC\omega_b - 1 = 0 \quad (3)$$

Here, letting $\omega_{b1} < \omega_{b2}$ and making the half width $B = f_{b2} - f_{b1}$, one can derive $Q = f_0/B$. This means that once the cut-off frequency f_b is obtained, the Q -value can be calculated using the relationship with the resonant frequency. However, even if the resonant frequency of the ATS beacon is known, it is impossible to find the cut-off frequency beforehand at the on-board inspection by a probe train. Therefore, it is necessary to sweep the frequencies around a frequency in the neighborhood of the resonant frequency and determine the Q -value from the changes of circuit current, and this method is discussed here. In general, the flowing current



decreases if the frequency is changed from the resonant frequency f_0 . Let I_d be the current value this time and I_0 be the current value at resonance. Then the following relationship is obtained between Q , I_d and I_0 .

$$Q = \frac{f_0}{2\Delta f} \sqrt{\left(\frac{I_0}{I_d}\right)^2 - 1} \quad (4)$$

Therefore, the Q -value can be calculated by measuring the circuit current when the frequency is changed. Next, on-board measurement is discussed.

2.4 The principle of Q -value measurement

The discussion in the previous section holds true if the current flowing in the beacon is observable. However, the beacon current cannot be observed from on board, so it is necessary to derive a theory to measure the current through the on-board pick-up coil or a pickup as an alternative to the on-board pick-up coil. If the on-board pick-up coil system is expressed as the equivalent circuit in Figure 2(a), the circuit formed when coupled with the beacon is shown in Figure 2(b). Letting M be the coefficient of mutual induction, the following simultaneous equations Equation (5) hold true.

$$\begin{cases} E_1 = j\omega L_1 I_1 + j\omega M I_2 \\ 0 = j\omega M I_1 + \left(j\omega L + r + \frac{1}{j\omega C} \right) I_2 \end{cases} \quad (5)$$

Solving these simultaneous equations,

$$\frac{I_1}{I_2} = \frac{r + j \left(\omega L - \frac{1}{\omega C} \right)}{-j\omega M} \quad (6)$$

Then, the current in the beacon can be calculated from the on-board current. However, L , C and M are constants of the beacon, so they are not observable from on board. Here, assuming that the on-board current is oscillating at the same frequency as the resonant frequency of the beacon,

$$\frac{I_{01}}{I_{02}} = \frac{r}{-j\omega M} \quad (7)$$

is obtained. On the other hand, assuming that the on-board current is oscillating at a frequency near the cut-off frequency,

$$\frac{I_{d1}}{I_{d2}} = \frac{r + j \left(\omega L - \frac{1}{\omega C} \right)}{-j\omega M} = \frac{r(1 + j)}{-j\omega M} \quad (8)$$



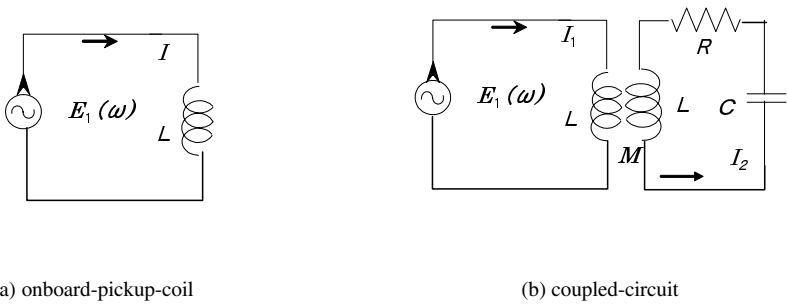


Figure 2: Equivalent circuit of ATS.

The current value is found using Equations (7) and (8), and the following expression is obtained.

$$\frac{I_{02}}{I_{d2}} = (1 + j) \frac{I_{01}}{I_{d1}} \quad (9)$$

Equation (10) can be obtained by substituting the absolute value of the expression into Equation (4).

$$Q = \frac{f_0}{2\Delta f} \sqrt{\left(\frac{I_0}{I_d}\right)^2 - 1} = \frac{f_0}{2\Delta f} \sqrt{2 \left(\frac{I_{01}}{I_{d1}}\right)^2 - 1} \quad (10)$$

By using Equation (10), it becomes possible to measure the Q -value of the ATS beacon from a probe train. That is, let I_{01} be the current value when the frequency of the on-board power is made equal to the resonant frequency of the ATS beacon and I_{d1} be the current value when the frequency is varied through the half-width. Then the Q -value can be calculated if these current values are found. As is obvious from Equation (10), this principle shows that inspection is possible irrespective of the value of the coefficient M of mutual induction between the on-board pick-up coil and beacon, or the value of L of the on-board pick-up coil, so it is valuable as a radical solution to the errors and measurement variations due to the difference in height of on-board pick-up coils, which has been a problem in conventional inspection cars.

2.5 Measurement principle verification test and results

To verify the derived measurement principle, the authors fabricated a beacon and on-board pick-up coil on a trial basis and conducted a test according to the following procedure.

1. Fabrication of simulated beacon

The authors fabricated the coil made of an LCR series circuit shown in Figure 3 as a simulated beacon, and found the impedances and resonant



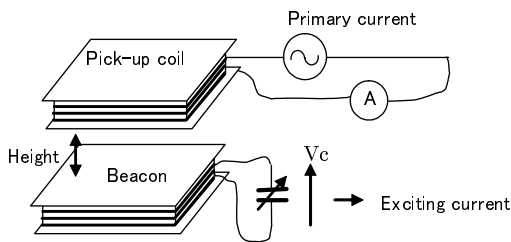


Figure 3: Test set of beacon and pick-up coil.

and cut-off frequencies for different capacitance values and calculated the Q -value for each.

2. Fabrication of on-board pick-up coil

The authors fabricated an pick-up coil made of the on-board coil and found the inductance and resistance values of the on-board coil.

3. Coupling test

The authors coupled the beacon and on-board pick-up coil and measured the change of primary and secondary current at that time at frequencies near the resonant and cut-off frequencies of the beacon. The same test was conducted with different coupling distances between the beacon and pick-up coil.

4. Test with different shapes of beacon

To elongate the coupling time when a train is running over, an oblong on-board pick-up coil was fabricated and a similar test was conducted using it.

5. Validation of test results

The validity of the theory was verified from the test results.

The test was conducted at various resonant frequencies and various distances between the beacon and on-board pick-up coil, and the results are shown in Figure 4. As can be seen from the test results, the developed technology is not influenced by variations of distance between the beacon and on-board pick-up coil, so it can ensure stable measurement.

2.6 Implementation of the Q -value inspection system

In the measuring method by this technology, it is necessary to excite the ATS beacon at frequencies near the resonant and cut-off frequencies of the beacon. On the other hand, the resonant frequency of a beacon differs according to the signal indications. To transmit a current wave of a conceivable resonant frequency and its cut-off frequencies to the beacon via the on-board pick-up coil, the method of sweeping the frequencies in the vicinity of the resonant frequency is conceivable. However, the time of coupling between the beacon and on-board pick-up coil is just 8 ms for a fast train. Thus, this method is not practical because of the severe time limit.



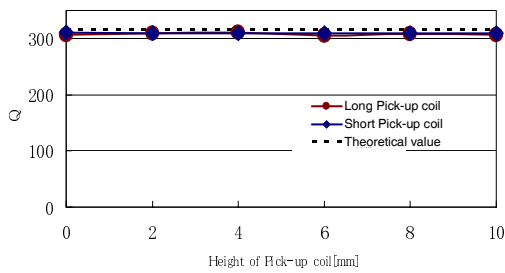


Figure 4: Effects for the alteration of pick-up coil's height.

Then, it was decided to prepare a composite probe signal containing signals of all the previously assumed frequencies and use it to excite the on-board pick-up coil. When the pick-up coil in which this composite probe signal is flowing passes over a beacon, the pick-up coil current changes. This current change is subjected to FFT analysis to extract the level of each frequency component and determine the resonance frequency of the beacon. Furthermore, the frequencies at which the level is 70% of the level of the resonant frequency are defined as cut-off frequencies, and the Q -value is calculated from the frequency bandwidth determined in this way.

The discussion thus far has proceeded on the assumption that the resonant frequency of the beacon is not deviated. In practice, however, the effect of deviations of the resonant frequency of the beacon must be taken into account. Thus, a test was conducted to see in what kind of error the deviations of the resonant frequency of the beacon manifest themselves.

The authors prepared a beacon having a resonant frequency deviated 0.2 kHz with respect to the reference resonant frequency of 130 kHz and tested it on the actual field. The half width for the resonant frequency is 0.6 kHz. The measurement results are shown in Figure 5. The resonant frequency of 130.2 kHz of the beacon was correctly recognized from on board. Cut-off frequencies were estimated from this and measurements were made. It was found that the error would fall within the range of several percent. If, on the other hand, the deviation of the resonant frequency is not recognized and a solution is derived without correction, the error amounts to about 10%, or reduces to the conventional level. For accurate implementation, therefore, it is important to prepare a composite probe signal comprising many frequency components.

For existing ATS beacons, the time of coupling with the on-board pick-up coil is about 8 ms if the train is running at high speed. Therefore, even if a composite probe signal comprising many frequency components is prepared, a fast real-time spectrum analyzer is needed to observe the level of each frequency component within 8 ms. This is not suitable as a probe train facility in view of cost and space. Thus, it was decided only to record the data once in the time domain while a train is passing over the beacon, using the on-board location fixing function. This recorded data is analyzed off-line to obtain the Q -value of each beacon.



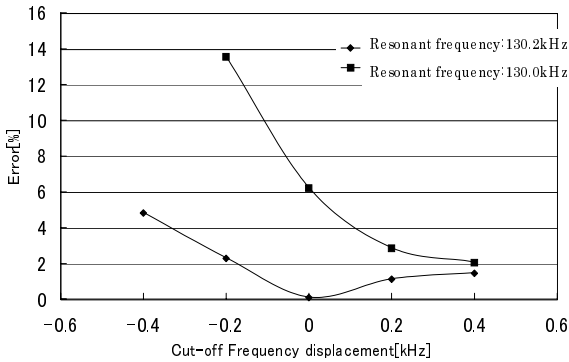


Figure 5: Measurement results of cut-off frequency displacement.

3 Track circuit measurement technology of probe train

Among signaling devices, the track circuit is a device to detect a train and an important element of the signaling equipment. The two-element three-position track circuit is a kind of track circuit and performs train detection using a track relay, which operates according to the relationship between the level and phase of the track circuit signal at the signal receiving terminal. The level and phase change if the track condition changes; that is, without appropriate maintenance, the track relay may detect a train erroneously although there is no train on the track. In such cases, train service becomes impossible. To prevent such a situation, periodic field maintenance takes place; however, it requires a large maintenance staff. As a measure, the field devices are placed in a centralized manner in a machine room in the station yard, and they are controlled there. However, this method incurs construction costs, so it is now restricted to the prime lines.

For measurement of the track circuit, the probe train implements automatic measurement management of the level and phase of the track circuit. For this purpose, a signal receiver is installed on board to take in the track circuit signal. In addition, an on-board signal processor finds the phase difference between the reference phase at the signal receiving terminal of the track circuit and the transmission terminal phase at the transmission terminal, and this phase difference is used as maintenance information for the track circuit. The method of measuring the track circuit from the probe train is described below.

3.1 Track circuit reception level measurement techniques

Reception level management of the track circuit is regarded as an important measurement item for maintenance. In track circuit measurement from the probe train, the track circuit reception level is measured by the on-board signal receiver, and the trend of level changes is used as maintenance data. In track circuit measurement from the probe train, therefore, the train location is fixed to identify the track circuit boundary. On the other hand, the on-board reception level is also influenced



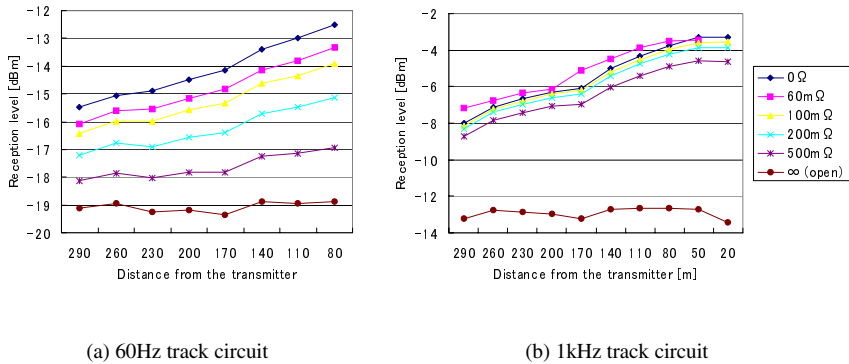


Figure 6: Reception level for each short-circuit resistance at different frequencies of track circuit.

by the short-circuit resistance of the train, so its examination is indispensable.

The reception level change characteristics due to the short-circuit resistance of the train were measured for each of the frequencies of the track circuit. The results are shown in Figures 6(a) and (b). Figure 6(a) concerns a 60-Hz track circuit, and the influence of short-circuit resistance is remarkable. Therefore, if only the level at the signal receiving terminal of the track circuit is measured from on board, making some judgment based on that level alone is dangerous. This can be resolved by estimating the short-circuit resistance of the train backward from the reception level at a managed signal transmission terminal and making a correction based on that resistance. On the other hand, Figure 6(b) shows the results of measurement performed on a 1-kHz AF(Audio Frequency) track circuit. The short-circuit resistance of the train has little influence, so there is no problem if the measurement values are used without correction.

3.2 Track circuit phase measurement techniques

In maintenance of the track circuit, phase measurement is also an important item along with reception level measurement. However, it has been considered difficult to inspect phase from a single received signal. The authors take the phase of the track circuit signal received at the receiving terminal when the probe enters the section of the track circuit as the reference phase and measure the phase when the probe train proceeds and reaches the signal transmission terminal of the track circuit (transmission terminal phase). The difference of these phases is used as maintenance data. Therefore, the probe train performs location fixing, that is, it knows the entry to the section of a track circuit for certain from information about the boundary points of track circuits.

In an actual track circuit, elements other than the rails which affect the phase are incorporated, so phase management must be done including these elements. This means that the mere detection of phase changes on the track circuit signal received



by the on-board signal receiver is not sufficient. On the other hand, the technique proposed here assumes that when a track circuit is installed, phase adjustment should be complete and the optimum condition should be attained. Taking the phase difference this time as the reference, it is planned to attain the objective by managing the changes of phase difference due to aging deterioration.

4 Conclusion

The authors are developing a probe train, and this article introduces the measurement technology, mainly about the track circuit and ATS technology, of the signal facilities on board the train. This study is promoted as a part of the fundamental research promotion scheme “Fundamental Study on Renovation of Track Traffic System by Introduction of Probe Train Technology (Study on Signal Facility Assessment Technology by Probe Train and Study on Data Gathering Techniques)” in the transportation sector by the Japan Railway Construction, Transport and Technology Agency, an independent administrative agency.

Among the ATS systems which occupy an important position as railway safety equipment, the frequency shift ATS is greatly used, irrespective of whether on a JR or private railroad. In this study, a method using changes of the on-board pick-up coil current was examined as an inspection method independent of the coupling coefficient or geometrical positional relationship between the beacon and pick-up coil, and the feasibility of this method has been confirmed. For data gathering for track circuit maintenance and measurement, it was decided to measure the level at the transmission terminal of the track circuit and the signal level at the receiving terminal and changes of phase between these terminals. As a preliminary examination, a field test was conducted to know the influence of the short-circuit resistance of the train. The conclusion obtained is that the influence is larger in the low frequency region than in the high frequency region. It is planned to verify the effectiveness of these techniques by a test using an actual train in 2006 and incorporate them as inspection and measurement functions of the probe train.

References

- [1] Tsunashima, H. & et al., Adaptive estimation of vehicle and road state using interacting multiple model approach. *Proc. of the 9th Int. Mini Conf. On Vehicle System Dynamics, Identification and Anomalies*, Budapest, 2004.
- [2] Tsunashima, H. & et al., Dynamics of actively steered automated guideway transit vehicle. *Proc. of the 10th Int. Conf. On Automated People Movers*, Orlando, 2005.
- [3] Tsunashima, H. & et al., Fault detection of railway track by multi-resolution analysis. *Proc. of the 4th Int. Conf. On Wavelet Analysis and Its Applications*, Macau, 2005.



Fault detection of railway track by multi-resolution analysis

T. Kojima¹, H. Tsunashima² & A. Matsumoto³

¹*Graduate School of Nihon University, Japan*

²*College of Industrial Technology, Nihon University, Japan*

³*National Traffic Safety and Environment Laboratory, Japan*

Abstract

Conditions of track have conventionally been measured by using exclusive inspection vehicles. If the conditions of the track can be detected by attaching simple sensors such as accelerometers on board of commercial vehicles, more efficient maintenance of tracks would be possible. This paper describes rail corrugation detection from the vertical acceleration of a vehicle body by time-frequency analysis. An actual vehicle running tests on a commercial line was conducted, in which vertical acceleration of axle-boxes and a vehicle body were measured. In this paper we show the rail corrugation can be detected from the vertical acceleration of a vehicle body by multi-resolution analysis (MRA) using wavelet transform.

Keywords: railway, spectrum analysis, diagnostics, rail corrugation, fault detection, wavelet, multi-resolution analysis.

1 Introduction

Faults of railway tracks not only cause deterioration in ride quality but also generate noise and have the potential to cause serious accidents. The track faults have conventionally been measured using exclusive track inspection vehicles. However, this method is costly and is therefore not widely employed on local and sub-main lines. It also has a problem in that very frequent inspections cannot be made if schedules are overcrowded, even on priority lines.

If faults can be detected by attaching simple sensors such as accelerometers to commercial vehicles, more efficient maintenance of tracks would be possible. Such vehicles are called “probe-vehicles”. In order to put “probe-vehicles” to



practical use, estimation of the vehicle's state, by pulling out the necessary information from among the information on track irregularities and the measurement noises of the sensors, is required.

This paper proposes the methods to detect a rail corrugation by using a wavelet transform from the acceleration measured in a vehicle body on board.

2 Measurement test with actual vehicle

2.1 Rail corrugation

Rail corrugation is minor wear and has a wave height of 0.1 mm and a wavelength of 5–10 cm occurring on the rails; it generates both damage to rails and noise [3–6]. It occurs mainly in the inner rails in steep curves such as shown in fig. 1



Figure 1: Example of rail corrugation.

2.2 Measurement test

Vertical and lateral accelerations of the vehicle body and the axle-box were measured with accelerometers, and exterior noise was also measured with microphone at the one of the main line in Japan (fig. 2). Focusing on the vehicle body acceleration as the object of the analysis, we attempted to extract the corrugation information.

2.3 Test results

Fig. 3 shows the measurement result of the curved section with significant corrugation. This is the measurement result for traveling a curve with a radius of 202 m at a constant speed of 38 km/h. A comparison of vertical acceleration of the right and left axle-boxes, shown in figs. 3(a) and figs. 3(b), respectively,



shows that the left axle-box, i.e. the inner-rail side, vibrates more strongly. This is a characteristic of corrugation in steep curves and confirms the occurrence of corrugation on an inner-side rail.

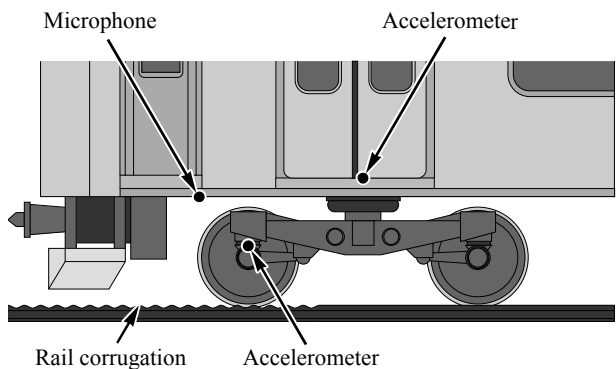


Figure 2: Position of sensors attached.

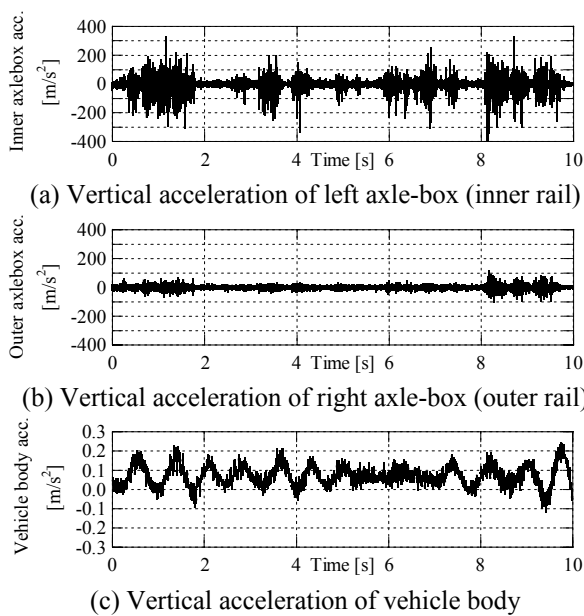


Figure 3: Measurement results of curved section with corrugation.

In order to show a comparison with a case where there is no corrugation, the measurement result of the straight section is shown in fig. 4. The travel speed was 38 km/h, as in the case for the curved section. The vertical accelerations of the both right and left axle-boxes are extremely small in comparison with those measured in the curved section shown in fig. 3. A difference in the accelerations



due to presence of corrugation is noticeable on the axle-boxes, particularly on the inner-rail side. The corrugation of this example could possibly be detected comparatively easily by methods such as threshold-value treatment.

However, when converting a commercial vehicle into a “probe-vehicle” by attaching sensors, the introduction can be made more easily by attaching sensors at locations as easily accessible as possible. This means that if track faults can be detected by accelerometers equipped in a vehicle cabin, fault detection of rails can be made without modifying a vehicle, leading to easier introduction.

The results for the curved section (fig. 3(c)) and the straight section (fig. 4 (c)) show that the vertical accelerations of the vehicle body measured on the floor of a cabin are greatly influenced by first resonance frequency of the vehicle body and thus, vibration due to the corrugation decays. These results indicate that detection of corrugation from the vehicle body accelerations requires special signal processing to extract the signal due to the corrugation.

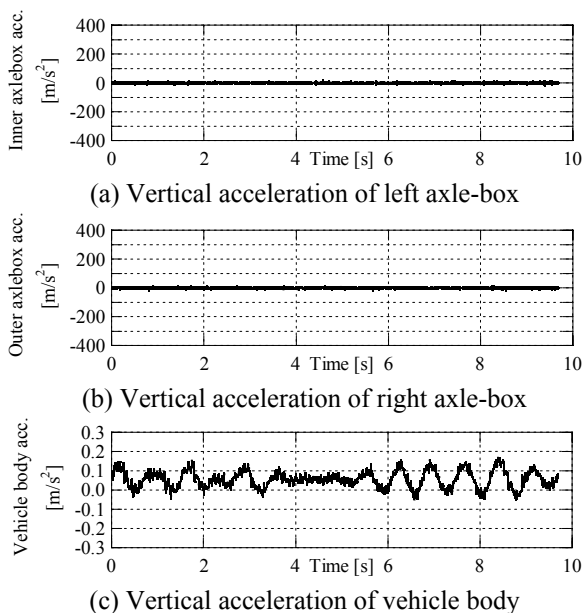


Figure 4: Measurement results of straight section without corrugation.

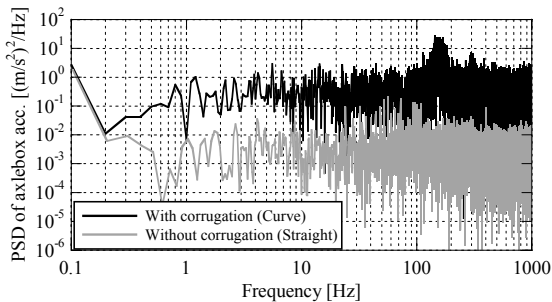
3 Detection of rail corrugation

3.1 Comparison through frequency analysis

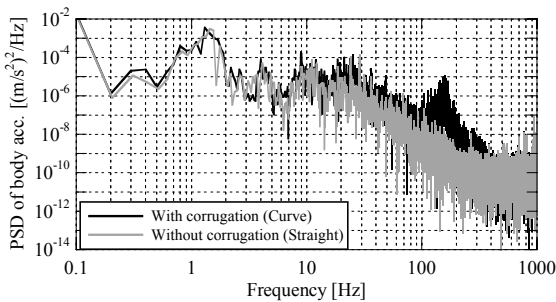
Since corrugation is a periodic on the rail surface, it was possible that the use of a frequency analysis would be able to extract its characteristics. PSDs were obtained from the data on the sections with and without corrugation (fig. 3 and fig. 4, respectively), and the results are shown in fig. 5. Fig. 5(a) shows the PSD



of the vertical acceleration of the axle-box of the inner-rail side. The curved section with corrugation has a higher level than the straight section over the entire frequency range and has a peak at approximately 160 Hz. Since the travel speed was 38 km/h, a wavelength of 6.9 m can be obtained from the peak frequency. This agrees with the general wavelength range of corrugation, and therefore the peak is considered to be caused by corrugation.



(a) PSD of acceleration of inner axle-box



(b) PSD of acceleration of vehicle body

Figure 5: PSD with and without corrugation.

Fig. 5(b) shows the PSD of the vertical acceleration of the vehicle body. The peak at 1–2 Hz is considered to represent the natural frequency of pitch of the vehicle body. In the curved section with corrugation, a peak is observed at approximately 160 Hz, as was found on the axle-box, indicating that vibration caused by corrugation is included in the vehicle body vibration. The presence of corrugation can be recognized through a frequency analysis. However, identification of the actual location of its occurrence requires a time-frequency analysis.



3.2 Time-frequency analysis using wavelet transform

3.2.1 Wavelet transform

Fourier transform is frequently employed for extracting frequency information from time history. However, it has the problem of losing time information that is important for determining the locations of faults. Although a windowed Fourier transform can solve this issue, it is not fit for detecting unknown signals. The result depends on size of the window. Small window cannot see low frequency. Large window blurs the rapid changes to the whole of the window.

Wavelet transform [2] is the expansion of Fourier transform. In wavelet transform, a signal $S(t)$ is described by shifting and scaling a small wave, $\psi(t)$, called a mother wavelet. The signal is analyzed on the basis of the mother wavelet transform. A continuous wavelet transform (CWT) is given by:

$$(W_{\psi}S)(a,b) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{a}} \overline{\psi\left(\frac{t-b}{a}\right)} S(t) dt \quad (1)$$

$\psi((t-a)/b)$ shifts the time (phase) by b and makes the frequency $1/a$. This continuous wavelet transform, however, requires a considerable amount of computation, and information is redundant. Therefore, by making $a=2^m$ and $b=2^m n$ in eqn (1), we obtain:

$$D_{m,n} = \int_{-\infty}^{\infty} S(t) \psi_{m,n}(t) dt \quad (2)$$

where:

$$\psi_{m,n}(t) = 2^{-m/2} \psi(2^{-m}t - n) \quad (3)$$

which is called a discrete wavelet transform (DWT). This discrete wavelet transform is characterized by more efficient translation of signals.

3.2.2 Multi-resolution analysis

Multi-resolution analysis (MRA) decomposes a signal into a number of components at different resolution by using the discrete wavelet transform. A time history $S(t)$ is decomposed into some detailed (high-frequency) components and an approximated (low-frequency) component. The original signal $S(t)$ given by:

$$S(t) = S_{m0}(t) + \sum_{m=-\infty}^{m0} d_m(t) \quad (4)$$

where d_m are detailed components:

$$d_m(t) = \sum_{n=-\infty}^{\infty} D_{m,n} \psi_{m,n}(t) \quad (5)$$



In the wavelet transform, the choice of a mother wavelet $\psi(t)$ is important. We employed *Daubechies* wavelet [1], which is orthonormal base and compactly supported wavelet. The vanishing moments of *Daubechies* wavelet can be changed by index N . We decided to use a relatively high-order generating index, $N = 7$ (fig. 6).

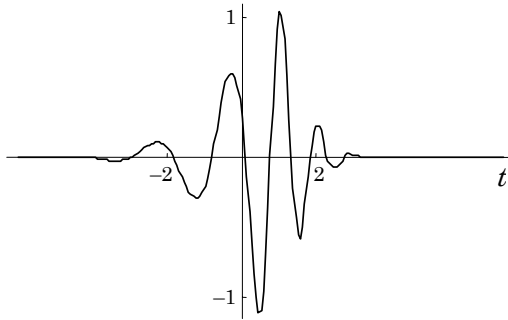


Figure 6: Daubechies wavelet ($N=7$).

3.2.3 Detection of corrugation through multi-resolution analysis

We propose a multi-resolution analysis to the vertical acceleration of a vehicle body, as measured on the vehicle cabin, in order to introduce a corrugation detection method in existing commercial vehicles.

The results of the multi-resolution analysis of the vertical accelerations in the curved and the straight section (fig. 3(c) and fig. 4(c), respectively) are respectively shown in fig. 7 and fig. 8. Because the sampling frequency was 2 kHz, d_1 , d_2 , d_3 , d_4 , and a_4 correspond to approximately 500–1 k, 250–500, 125–250, 62.5–125 Hz, and frequencies not greater than 62.5 Hz, respectively. A comparison between fig. 7 and fig. 8 shows that the components d_2 through d_4 for the case with corrugation, fig. 7, have larger amplitudes. Component d_3 , in particular, which includes the frequency, 160 Hz, for corrugation, shows conspicuous acceleration, which is of a waveform similar to that for the axle-box shown in fig. 3(a). This result indicates that the vibration component due to corrugation can be extracted from the acceleration of a vehicle body by using a multi-resolution analysis.

The result of the analysis with a *Daubechies*' generating index N of 7 is shown here. As a result of analyses with varied N 's, we attained good detectability with $N \geq 4$ for corrugation.

4 Conclusions

With the aim of equipping a commercial vehicle with lowcost sensors and thus putting into operation a "probe-vehicle" that would enable track inspection, we studied methods of detecting rail corrugation from data measured on a vehicle. We obtained the following conclusions:



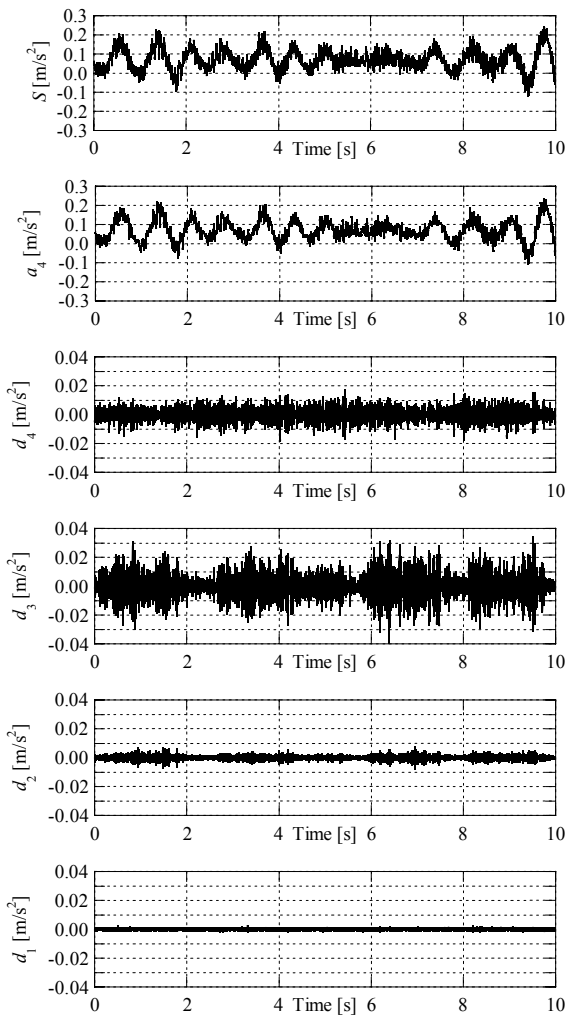


Figure 7: MRA of acceleration of vehicle body with corrugation (curved section).

- (1) While the vehicle was traveling on rail corrugation during the field test clear vertical vibration occurred on the axle-box on the inner-rail side. However, detection of corrugation from the amplitude of the acceleration of the vehicle body was difficult because it includes large amount first resonance vertical vibration mode.
- (2) Frequency analyses revealed peaks in all spectra of acceleration and the noises of the axle-boxes and vehicle body in the section with corrugation, and it was clear that corrugation generated periodic vibration.



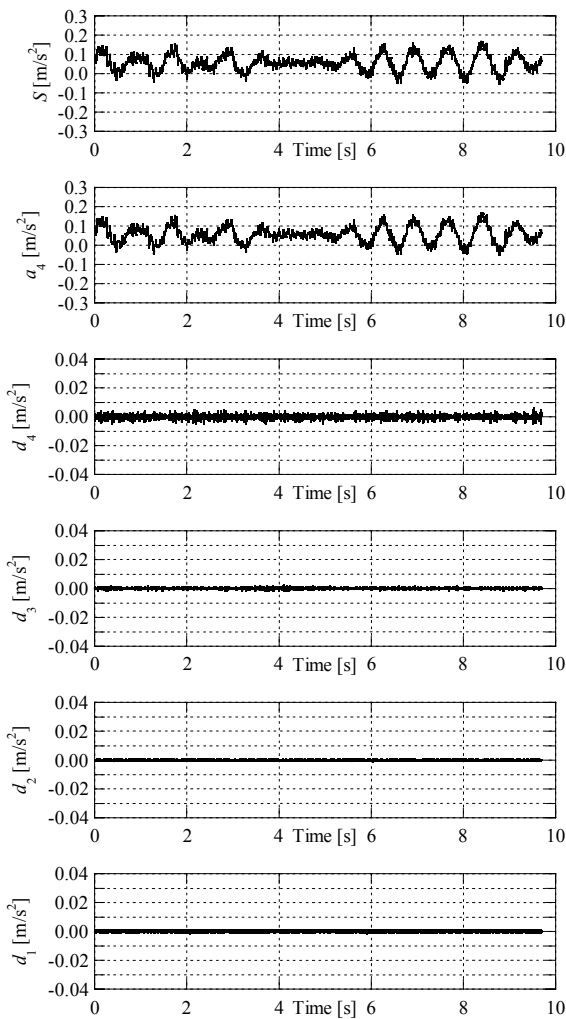


Figure 8: MRA of acceleration of vehicle body without corrugation (straight section).

- (3) Multi-resolution analyses were carried out using wavelet transform, and vibration component for corrugation was extracted from detailed component (d_3) including frequency of corrugation. This verified that corrugation could be detected from the acceleration of a vehicle body.

Here, we have focused on the study of corrugation as one track fault. The application of these methods to other track irregularities is conceivable as a future expansion of this research. Since the detectability of discontinuity is part of the character of the wavelet transform, we expect that it will be possible to detect impulse-like signals such as abnormalities of rail joints.



This study was supported by the Program for Promoting Fundamental Transport Technology Research from the Japan Railway Construction, Transport and Technology Agency (JRRTT).

References

- [1] Daubechies, I., Orthonormal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics*, **41(7)**, pp. 909–996, 1988
- [2] Daubechies, I., Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series In Applied Mathematics: *Society for Industrial and Applied Mathematics*, **(61)**, 1992
- [3] Ishida, M., Moto, T. & Takikawa, M., The effect of lateral creepage force on rail corrugation on low rail at sharp curves, *Wear*, **253(1)**, pp. 172–177, 2002
- [4] Matsumoto, A., Sato, Y., Ono, H., Tanimoto, M., Oka, Y. & Miyauchi, E., Formation mechanism and countermeasures of rail corrugation on curved track, *Wear*, **253(1)**, pp. 178–184, 2002
- [5] Suda, Y., Hanawa, M., Okumura, M. & Iwasa, T., Study on rail corrugation in sharp curves of commuter line, *Wear*, **253(1)**, pp. 193–198, 2002
- [6] Suda, Y., Komine, H., Iwasa, T. & Terumichi, Y., Experimental study on mechanism of rail corrugation using corrugation simulator, *Wear*, **253(1)**, pp. 162–171, 2002



Research on the onboard auto-test system for track circuit compensating capacitors

H. Zhao, Y. Liu & B. Liu

*School of Electronics and Information Engineering,
Beijing Jiaotong University, China*

Abstract

A kind of onboard auto-test system was proposed so as to detect the track circuit compensating capacitors in real time. The system model and design were based on the resonance current detection method. The data acquisition card generated the actuating sine signal first, which was amplified and coupled onto the rail. Two wheels of the test vehicle, rail lines between and the capacitor connected between the rails formed a closed LC loop. If the actuating signal frequency coincides with the inherent resonance frequency of the loop, a reactive current will be produced within the loop. After the current is sampled and further processed, the capacitor value can be estimated. The system principle and architecture are described in detail. Experiments show the invalid capacitors can be efficiently detected and located in real time.

Keywords: resonance current detection method, compensating capacitor, onboard auto-test system, track circuit, data acquisition.

1 Introduction

Audio frequency jointless track circuits, i.e. UM71 and ZPW-2000A, are now widely used in the railway autoblock systems in China. As for ZPW-2000A, rail lines act as the transmission media for dedicated FSK signal which carried the vital autoblock information used for train control. Modeled as a kind of distributed transmission line with inductive characteristic impedance, rail lines are not a good transmission media for audio frequency, unfortunately. In order to weaken the negative effect of the inductive rail and prolong the transmission length of track circuit, compensating capacitors are evenly spaced and installed between the rail lines, as shown in Fig. 1.



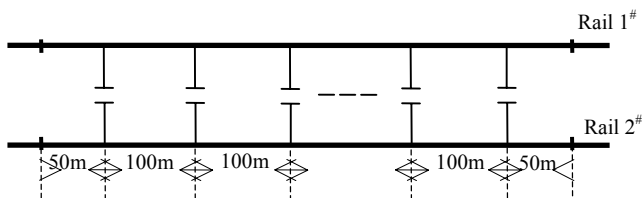


Figure 1: Compensating capacitor installation between rail lines.

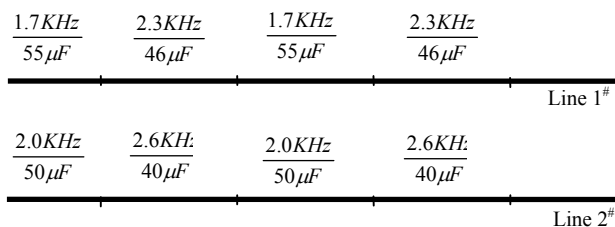


Figure 2: Compensating capacitor and carrier frequency configuration for Line 1# and Line 2# in Chinese railway.

For double lines, Line 1# and Line 2# use FSK modulated signal with different pair of carrier frequency. Besides that, the frequency is configured alternatively, as shown in Fig. 2. Hence, different compensating capacitor values should be used and configured correspondingly to match the different carrier frequency [1].

However, the evenly spaced compensating capacitors bring about another boring problem, i.e. the normal detection and maintenance of these capacitors is rather costly. The need for efficient onboard auto-test system is urgent.

Ref. [2] proposed a kind of compensation capacitor detection method based on wavelet analysis method by making use the inducing voltage curve received by cab signal receiving coils. But, this method can only judge whether the capacitor is fully invalid or not. Furthermore, it is only effective when the track circuit ballast impedance is approximately 1~3 ohm.km. As the ballast impedance is often influenced by weather, geography conditions and other environmental factors and varies from 1~300 ohm.km, the method is hardly to be used to guide maintenance.

At present, test vehicle has been the important measure for the online detection and maintenance of railway autoblock equipment. In this paper, a new kind of compensating capacitor auto-test system equipped on the test vehicle is put forward based on the resonance current detection method.



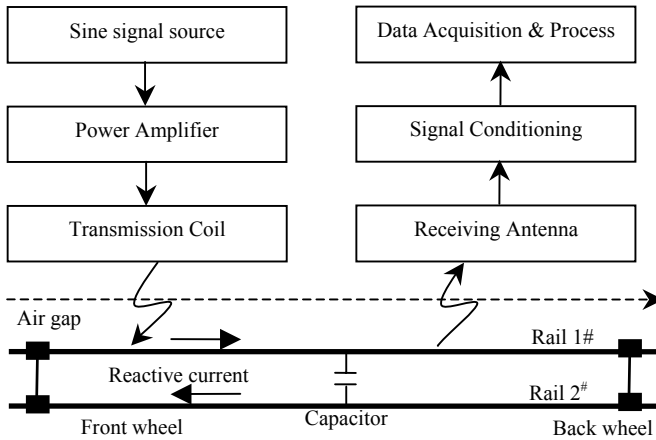


Figure 3: Principle of the resonance current detection method.

2 Principle and system architecture

2.1 Resonance current detection method

As shown in Fig. 3, a sine signal generator and a power amplifier are fixed on the test vehicle. Actuating sine signal is generated by the signal generator, amplified by the power amplifier, and then transmitted to the transmission coil installed under the vehicle. As the vehicle wheels and the rail lines between formed a closed loop, the magnetic field caused by the actuating signal will surely produce reactive current within the rail lines even though the current magnitude is relatively small. While the test vehicle passes the rail lines equipped with compensating capacitor connected in between, the above closed loop changes to an equivalent LC loop. Supposed that the actuating signal frequency is close to or equal to the resonance frequency of the LC loop, the reactive current will be enlarged to great extent.

The receiving antenna installed above the rail sense the current, transform it to proportional reactive voltage, and then transmit it to the signal conditioning unit. The data acquisition and process unit located backward analyzes the magnitude alteration of the reactive voltage and makes judgment whether the vehicle has passed a compensating capacitor. Capacitor value can be further calculated as the whole system is carefully scaled.

2.2 System architecture

Fig. 4 shows the installation position of dedicated test equipment under the test vehicle. A pair of transmission coil (TC1 and TC2) is fixed behind the front wheels, whereas another pair of receiving antenna (RA1 and RA2) is fixed before the back wheels. Speed sensor installed on one front wheel is used for train positioning.



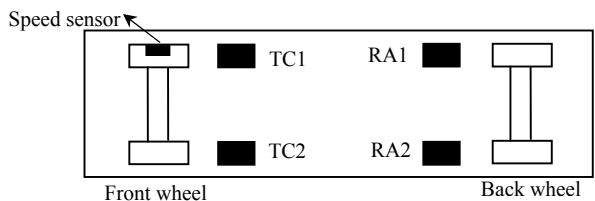


Figure 4: Installation position of dedicated test equipment under the test vehicle.

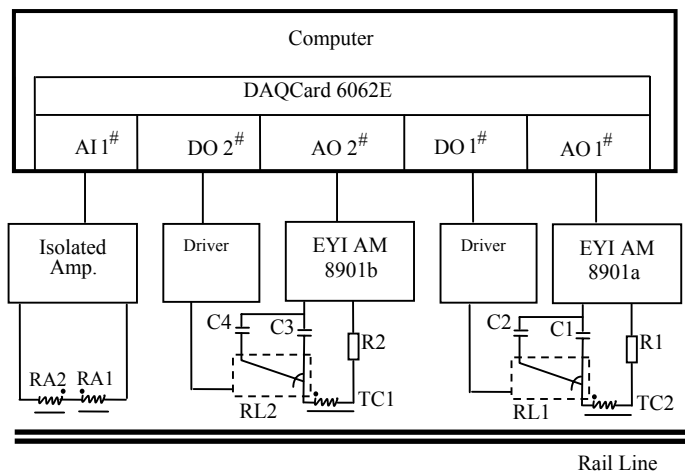


Figure 5: Auto-test system architecture. Note: R1, R2: 2Ω/50w; C1: 51 nF; C2: 12 nF; C3: 72 nF; C4: nF; TC1, TC2: 10 mH, Q = 21; RA1, RA2: 63 mH, Q = 5.

Fig. 5 shows the architecture of the auto-test system. Computer and the data acquisition card of NI DAQCard 6062E are the key components of the system. The analog output channels AO1[#] and AO2[#] generate actuating sine signal, which is amplified by power amplifier EYI Am8901 and transmitted to the transmission coil TC1 and TC2 respectively. Within each AO channel, resistors R1 and R2 are used to limit the loop current; whereas capacitors C1~C4 are used to counteract the inductive effect of transmission coil TC1 and TC2, improving the output current.

The signal generated by AO1[#] is used to actuate the equivalent LC loop with compensating capacitor value of 40e-6 F or 46e-6 F; while the signal generated by AO2[#] is used to actuate the LC loop with capacitor value of 50e-6 F or 55e-6 F. The two AO channels work at the same time, facilitating the simultaneous capacitor detection of 40e-6 F and 50e-6 F for Line 1[#] and 46e-6 F and 55e-6 F for Line 2[#].

The digital output channels DO1[#] and DO2[#] operate the power relays RL1 and RL2, making C1 and C3 connect with C2 and C4 wherever needed.



The analog input channel AI1[#] acquires the reactive voltage received by the receiving antenna RA1 and RA2 and isolated by ISO 124. RA1 and RA2 are connected each other in certain way so as to eliminate the interference caused by electrified traction current in electrified railway.

2.3 Software implementation

System software was developed under NI LabWindows /CVI, a kind of efficient C API for virtual instrumentation. System software functions are described as follows:

- 1) Generating actuating sine signal through AO channel, controlling the power relay through DO channel and selecting appropriate capacitor value (C1~C4)
- 2) Acquiring the reactive voltage through AI channel, filtering the s Signal in frequency domain and calculating the rms value of the reactive voltage
- 3) Comparing the reactive voltage with the reference value, estimating the compensating capacitor value by polynomial curve fit method which is realize by the “*PolyFit()*” function in LabWindows/CVI.
- 4) Combining with the train positioning system, give the coordination of the capacitor detected.

3 Experiment and data analysis

A series of experiment are carried out while the auto-test system is setting up. The aim is to determine some obscure parameters and verify the validity of the resonance current detection method.

3.1 Inherent resonance frequency determination experiment

Although the compensating capacitor value and the rail impedance between the test vehicle wheels are predetermined, there still exist many obscure factors, including the wheel impedance, contact impedance between wheel and rail, the connecting impedance of capacitor with the rail, etc. Experiment should be done to determine the inherent resonance frequency of the equivalent LC loop. The normal static measurement method for system frequency response feature is used in the experiment [3]. The result is shown in Table 1.

The actuating signal frequency generated by AO channels of DAQCard 6062E can be fully determined according to Table 1.

3.2 Capacitor value sensitivity experiment

Whether the reactive voltage is sensitive enough to the value change of the compensating capacitor within the equivalent LC loop is actually determined by the Q value of the LC loop itself. In the experiment, fixed rms value of actuating signal with specified frequency shown in Table 1 was firstly generated by Channel AO1[#]. The reactive voltage was then recorded as the capacitor value is changed. The result is shown in Table 2.



Table 2 shows the reactive voltage is sensitive enough for the capacitor value measurement and the resonance current detection method is practical indeed.

3.3 Capacitor position sensitivity experiment

This experiment is used to check if the reactive voltage is sensitive enough to the position change of the compensating capacitor within the equivalent LC loop. The result is shown in Table 3.

Table 3 shows the largest reactive voltage appeared as the capacitor locates in center of the two wheel pairs. When the capacitor was moved outside the wheel pairs, the reactive voltage dropped rapidly. This shows the resonance current detection method is practical from another point of view.

Table 1: Result for inherent resonance frequency measurement.

Compensating Capacitor (F)	40 e-6	46 e-6	50 e-6	55 e-6
Inherent Resonance Freq. (Hz)	7050	6350	5950	5850

Table 2: Result for capacitor value sensitivity experiment.

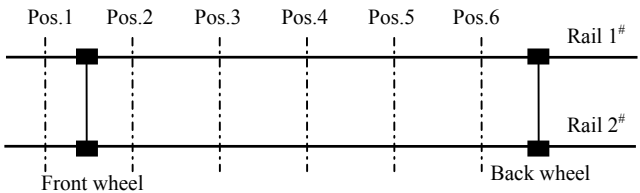
Actuating Signal Frequency(Hz)	Compensating capacitorF)	Reactive Voltage (mV)
7050	None	65.3
	21.4 e-6	98.7
	26.2 e-6	113.4
	33 e-6	133.8
	40 e-6	147.7
	46 e-6	90.0
6350	None	57.4
	21.4 e-6	77.9
	26.2 e-6	86.9
	33 e-6	103.0
	40 e-6	118.2
	46 e-6	135.3
5950	None	53.8
	26.2 e-6	74.7
	33 e-6	85.2
	40 e-6	96.1
	46 e-6	107.5
	50 e-6	117.8
5850	None	52.9
	33 e-6	80.2
	40 e-6	89.7
	46 e-6	102.6
	50 e-6	112.3
	55 e-6	112.6



Table 3: Result for capacitor position sensitivity experiment.

Position*	1	2	3	4	5	6
Reactive voltage (mV)	67.5	107.2	122.3	143.1	120	101.3

* The position is shown as in the following figure.



4 Conclusion

The onboard auto-test system of track circuit compensating capacitor is described in the paper. Conclusions are shown as follows:

- 1) Experiments show the resonance current detection method proposed in the paper is valid and efficient.
- 2) The compensating capacitor value can be estimated in real time, regardless the ballast impedance and other environmental factors.
- 3) Combining with the train positioning system, the coordination of the compensating capacitors under test can be provided.
- 4) Commercial data acquisition and virtual instrumentation techniques are applied in the auto-test system, ensuring the development efficiency and test accuracy.

References

- [1] H. Zhao, Y. Zhang, and X. Wang, "Study on Several Problems of Track Circuit with Compensating Capacitors," *Journal of Railways*, China, vol. 20, no. 4, pp. 77-81, August 1998.
- [2] L. Zhao, B. Liu, and X. Xu, "Using the Wavelet Analysis based Noise Canceling Technique to Detect The Integrity of The Compensating Capacitors of UM71 Track Circuit," *IEEE 2005 International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications Proceedings*, 3rd ed., China, vol. 2, pp.1263-1269 , August 2005.
- [3] H. Jiang, X. Sun, *Electronic Measurement*, Beijing: China Metrology Publishing House, 1988.



This page intentionally left blank

Monitoring wheel defects on a metro line: system description, analysis and results

M. Seco, E. Sanchez & J. Vinolas

CEIT and Tecnun (University of Navarra), Spain

Abstract

This paper presents a system developed to monitor wheel defects on metro vehicles in order to reduce annoyance and infrastructural damage due to noise and vibrations. Information about the wheel state and its evolution can be added to the mere mileage to improve the efficiency of the reprofiling decisions. In addition, sudden defects, that could potentially be severe, are detected and an alarm is raised to take the necessary actions.

Some other systems have been developed to monitor wheel state by means of measuring the vibrations or noises generated by vehicles; however, the case of a metro is somewhat distinct. Wheel thread defects on a metro vehicle are not necessarily related to overloads or rolling contact fatigue. Damage to infrastructure and noise emission generated by small defects are particularly important due to the high frequency of the service and the proximity of the track to residential buildings, especially in the case of an open-air section that runs through a neighbourhood as some sections of the Metrobilbao.

The system measures vibrations at a specific location on the rail as the vehicle passes over it, then analyzes these signals and stores the information in a specific database that can be easily queried from any computer on the intranet. The system has been installed on the Metrobilbao line and tested for three years. During this period, more than 8500 trains, or 272000 wheels, have passed over the accelerometers and corresponding data was stored in a database. 448 wheels threads were physically measured in order to check the accuracy of the system. This data has been used to fit the statistics for new or newly machined wheels and estimate the alarm threshold, and finally tuned in agreement with the maintenance workshop.

Keywords: wheelflat, monitoring, wheel/rail contact.



1 Introduction

The Railway industry is under increasing pressure due to the economic competition and requirements on comfort, on noise emission and safety for passenger and goods.

Simultaneously, a gain in speed and axle load results in increasing stresses in the wheel/rail contact, raising wear and maintenance costs. Little wear or small defects on rolling surfaces may result in overloading the wheels, bearings, rails and sleepers together with greater vibrations and acoustic emission. The most commonly reported wheel defects are known as flats and shells and it is estimated that \$90 million dollars are spent in the USA replacing 125000 defective wheels a year [1].

All of these problems are the reason for the growing interest in the Railway industry in preventive and predictive maintenance of infrastructures and rolling stock that has led to the development of different systems to detect wheel defects. Since the first "Wheel Impact Load Detector" (WILD) device was installed in the USA in 1983 [2], different systems and technologies have been proposed or developed [3-19]. However, correct quantification and monitoring are not definitely solved problems.

The generalized method to detect faulty wheels is based on the measurement of rail loads or vibrations, for wheel defects generate great impact loads, vibrations and high stresses on the rail. Different techniques have been used to measure the vibrations: accelerometers on the rail, strain gauges, load cells, etc. and a variety of algorithms have been developed to analyze measured signals.

Systems based on accelerometers are the Wheel Monitoring System (WMS) [8], the Wheel Impact Monitor (WIM) [10] and the one presented by Skarlatos et al. [18]. Other methods, [12, 13, 17], use the cepstrum function [20] to analyse the signals from piezoelectric cables and accelerometers.

The Wheel Condition Monitor (WCM) [19], and the TrackAlert® system [5] use both accelerometers and strain sensors installed on the rail while other systems use load cells or other equivalent devices to measure load peaks instead of accelerations [4, 6, 11].

Feng et al. [15] have developed a different device, based on the measurement of the vertical movement of the wheel flange by means of a rule attached to the rail. The principal limitation of the system is the relative low train speed required to work properly and its mechanical complexity.

Jodon, inc [11] propose a system based on microwave that measures the wheel profile and would be also capable of measuring wheel defects over the rolling band.

This article presents a low-cost monitoring system, based on accelerometers, developed by CEIT and Metrobilbao that was placed and successfully tested in the Metrobilbao railway system. Principal advantages of the system are precisely its low cost and simplicity.

This paper includes a system description on section 2, the signal processing and treatment on section 3, the experimental verification and alarm threshold on section 4 and conclusions on section 5.



2 System description

When an impact is applied to the rail, vibrations are propagated along the track decaying with distance. It is then necessary to know both the decay rate and the distance from the impact point to estimate the accuracy of the measurements.

Although extensive studies of the wave propagation along the rail have been published [21, 22], measured data have been used to simply estimate the specific decay rate of the registered vibrations as the flat impacts repetitively along the rail. An example of the registered signal corresponding to a Metrobilbao vehicle with a wheelflat is shown in Figure 1: Y-axis shows the RMS values of the acceleration in dB during the train passing by (X-axis). Impacts on the rail are due to a defect that affects the rolling contact point at every turn of the wheel.

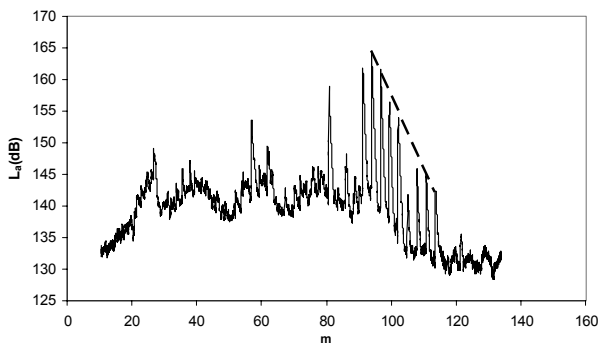


Figure 1: L_a (ref= $1\mu\text{m/s}^2$) of the acceleration signal for a Metrobilbao vehicle with a wheelflat in the rear last wheel. The slope of the fitted line is the spatial track decay rate for the impacts due to wheel defects.

A regression line can be fitted to the maximum of the different impact echoes shown in Figure 1: and the slope of the fitted line is thus the spatial decay of the excitation.

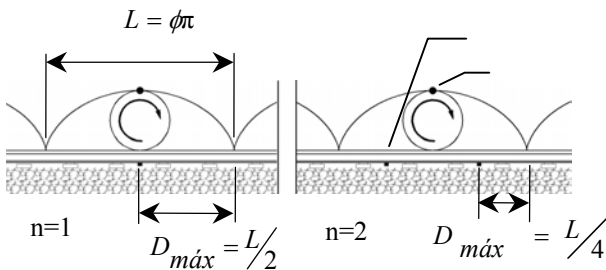


Figure 2: Defect paths and Dmax for 1 and 2 sensors on rail.

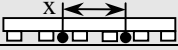
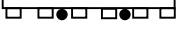
As the decay rate is rail dependent, the only way to achieve more precise measurements of the vibrations is to minimize the maximum distance (D_{max}) in which the excitation may occur because the bigger is the distance between the



impact and the measurement point, the bigger is the underestimation of the signal. The way to minimize $D_{\max}$ is to increase the number of sensors installed on the rail [10, 19]. For a single sensor scheme $D_{\max}$ is half the circumference length, with 2 sensors it is a quarter of the wheel circumference, etc (Table 1:).

Finally, 2 accelerometers have been installed at each measuring point looking for a half-way between costs and maximum admissible error.

Table 1: Different options for the accelerometers scheme.

n	Mounting Scheme		$D_{\max}$	Attenuation	
		m		ΔL_a (dB)	a/amax (%)
1		--	$L/2=1.32$	-1.584	83.3%
2		1.32	$L/4=0.66$	-0.79	91.3%
3		0.88	$L/6=0.44$	-0.53	94.1%
4		0.66	$L/8=0.33$	-0.4	95.5%

X: Distance between accelerometers. Where $x = \frac{\phi\pi}{n}$ with n = number of accelerometers

and ϕ = wheel diameter.

$D_{\max}$: Maximum distance between the measurement point and the impact point

ΔL_a : Theoretical decay in dB, where $\Delta L_a = DR \times D_{\max}$ and DR is the estimated Decay Rate.

a/amax: Maximum underestimation of the real excitation due to spatial decay.

The main components of the developed device are the accelerometer matrix under the rail, a wayside sonometer, signal transmission circuits, and the computer that controls the system, analyzes the signals and stores data. A couple of proximity detectors activate the system when a train approaches and are also used to estimate the speed of the train.

Figure 3: shows and schematic of the device installed close to a station in the Metrobilbao network. Notice that the equipments are spread into three zones:

- Control booth: where the computer and the acquisition electronics are placed.
- Bend zone: 4 accelerometers, a sonometer and a proximity detector.
- Straight zone: 4 accelerometers, and a proximity detector.

Some of the accelerometers have been installed on a bend because during turning the wheel-rail contact point moves along the wheel and rail profiles and this may allow one to detect different out of the rolling band defects.

A1 to A4 accelerometers on Figure 3: are far over 200m from the computer so a current transmitter circuits system has been installed to avoid noise and signal attenuation. Designed voltage-to-current converters transmit conditioners signals and current-to-voltage converters reverts the signals to the acquisition card inputs.



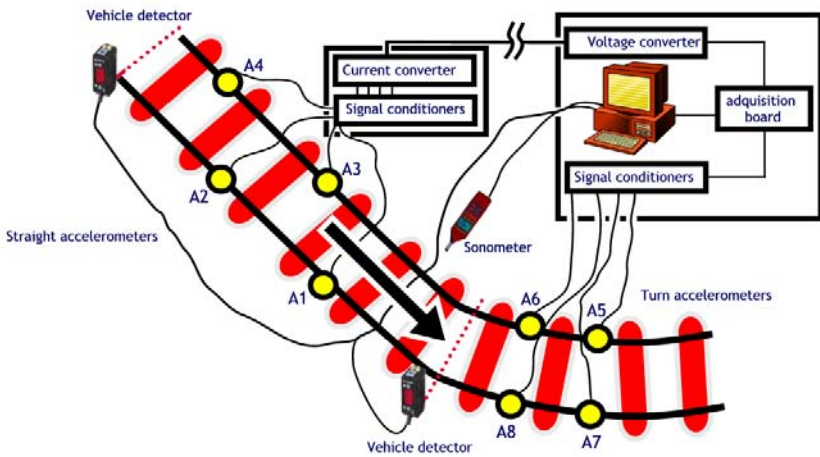


Figure 3: The system has been installed in the Metrobilbao network.

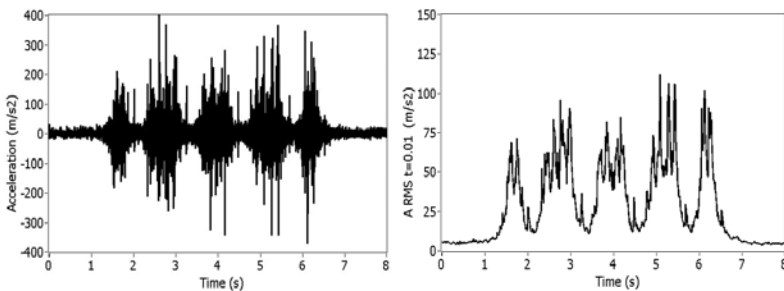


Figure 4: Rail vibrations during the train passing. Train without defective wheels. Left) original signal. Right) RMS integrated signal.

The sampling rate chosen for the signal acquisition is 2 kHz, since it was verified that the acceleration spectra of interest is below 1 kHz [17]. A sample for the registered rail acceleration pattern for a defect-free train is shown in Figure 4:. Obviously, the higher is the sampling rate, the better the signal but 2 kHz is high enough for monitoring necessities.

Accelerometers are fastened with a magnet and it has been used for it is simple, easy to install and move and it has remained successfully fixed during the system development.

3 Signal analysis and software

Acquired data is analyzed and stored and that makes post-processing possible because none of the data is deleted or modified during analysis allowing the to



alarm thresholds to be updated or to study particular trends, peculiarities or outstanding values without waiting for new data.

Some parameters, apart from the train identification, needed to carry the monitoring out, are the train speed and the train geometry.

The variable used for monitoring is the root mean square of the signal (Figure 4:).

Exhaustive tests have been carried out looking for the optimum value for the time step and finally it has been chosen in order to assure that faulty and non-faulty levels are clearly different and variability within the same state is minimal (Figure 5:).

The integrated values from each of the wheels for every vehicle transit are stored in a database. This database is the core of the system and different data requests are used for report warnings, defect evolution or instant wheels state. This configuration allows great flexibility in order to carry out different studies.

The system makes use of a couple of sensors within each of the rails to minimize the variability of the measures. Data from both sensors, corresponding to each of the wheels, are analysed but only the maximum values from both of the signals are stored on the database as corresponding to the level of each of the wheels.

The system raises the alarm when a defective wheel is detected but available working time and, especially, economical and wheel durability are, up to now, the leading parameters on the turning or changing wheels decisions. For this reason, the alarm threshold must include not only the mere defect importance but also the repeatedly machining costs in order to optimize the wheel life-cycle

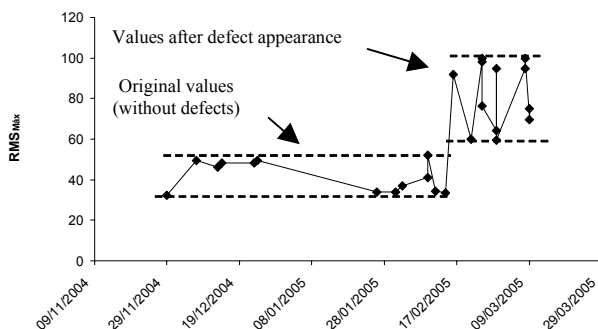


Figure 5: Evolution of the $RMS_{\Delta t=0.04}$ values with the apparition of a defect.

4 Alarm threshold tuning and experimental verification

Some studies have been published about the allowable defect magnitude in order to avoid rail damage. However, the problem in metro systems is more related to generated noise than to rail damage. In fact, great defects, when they occur, are currently quickly detected while a relative high noise level is particular annoying in open air tracks that are close to buildings and houses, even more with high



train frequency and if the service begins soon in the morning a finishes late at night.

The first approach to the alarm level tuning is to discern between ‘perfect’ and defective wheels. To do so, logarithms of the $RMS\Delta t=0,04$ values corresponding to new or just turned wheels are used to fit a reduced extreme values distribution as seen in next figure.

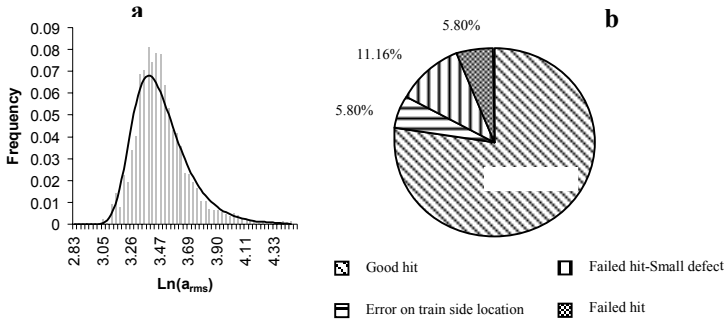


Figure 6: a) Experimental histogram of the logarithm of the acceleration for perfect or newly machined wheels ($n=2847$). Black line corresponds to the theoretical reduced extreme value distribution adjusted to the experimental values. b) Reliability of the device during machining process supervision ($n=448$).

Fitting is performed by means of maximum likelihood estimators and 99% confidence intervals are proposed to identify out of limit values that are considered to correspond with faulty wheels.

In order to involve maintenance department and workshop operators in the system developing, weekly alarms were reported as well as information about the state of the vehicle scheduled to be turned. This has allowed making use of the workshop experience to conform alarm thresholds to operative values.

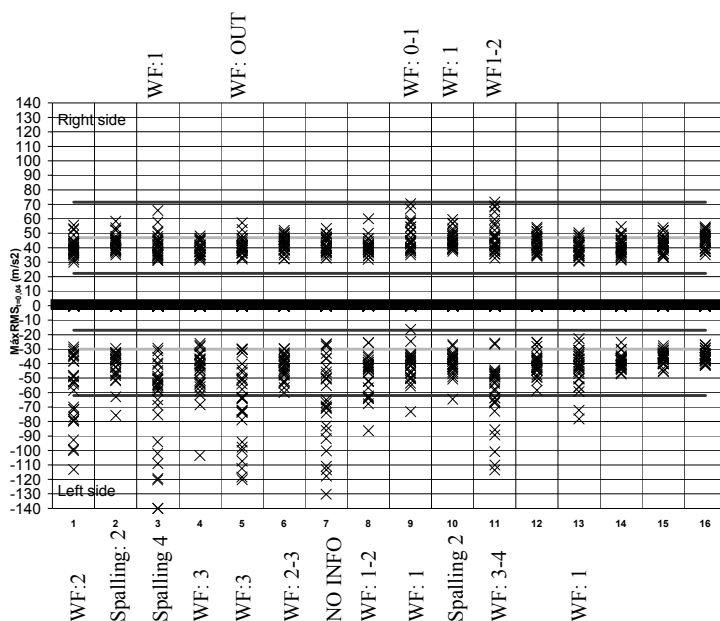
Tests included inspection of the turning of 448 wheels and 77% resulted in good agreement with the device estimation while 5.8% were considered to have failed. The rest of the failed estimations may be explained as follows:

A) The defect has been localized in the same axle but in the opposite wheel. Such mistake may be due to practical error of the workshop operator when reporting his observations.

B) Alarms with level 1 out of 4, which are not reported as having a defect may be caused because the device raises a low level alarm due to effects not related to wheel defects or it may be due to visual misdetection of a small defect during machining operations. Any way, these faults were not critical because the alarm level was low and it could be solved raising the alarm threshold level.

Figure 7: shows an example of the correlation between the monitoring based alarms and the defects found during the turning operation.





WF= wheel flat with level from 0 (no wheelflat) to 4

On the 5TH wheel on the right a small flat was reported to be out of the rolling band.

Figure 7: Comparison between data from the monitoring system (plot) and defect found during wheel re-profiling process (up and down).

The central graph shows vehicle monitoring information. The axis of abscissas shows the different wheel number of the monitored vehicle and the axis of ordinates measures the maximum acceleration values. The graph includes data from the right side wheels (up) and the left ones (bottom) and summarizes ten passages of the studied train. Horizontal lines delimit theoretic normal values and outstanding data are considered to correspond to faulty wheels. The figure also shows information reported from machining workshop, it includes results from visual inspection, machining incidents and other information.

5 Conclusions

Defects on the rolling band in train wheels are the cause of large impacts on the rail that generate noise and vibrations. With this work it has been proved that measuring these rail vibrations is a reliable method to determine the presence and magnitude of wheel defects. This paper presented a detailed study of the existing and currently developing technology in measuring techniques and analyzing algorithms. Also, it has been proved to be precise enough to use two accelerometers beneath each rail. Additionally, it is interesting to install accelerometers in different positions, straight line and turning zone because defects in different locations on the wheel surface are so detected.



Speed estimation is performed by means of the presence detector signals and adequate accuracy has been achieved. Speed accuracy is important in order to discard that measures corresponding to vehicles running with out of normal speed. In addition, vehicle speed is used during the signal analysis to correlate the signal section with its corresponding wheel.

In this study 448 wheels have been studied, comparing the predicted status with data from the machining process. Only 5,8% of the predicted status clearly failed while the other 5,8% of the predictions resulted in minimal defect misdetection.

The fully automatic operation of the system is under development by means of the automatic identification of the vehicle and the connection of the system with the data network of Metrobilbao.

Acknowledgements

Developing the system would have been an impossible task without the support of Metrobilbao. They have always trusted that the system would finally work so they have financed the work and have been at our beck and call to use their installations and rolling stock to carry out measurements, many times at night.

References

- [1] R.V. Dukkipati and R. Dong, Impact loads due to wheel flats and shells, *Vehicle System Dynamics*, 31(1),pp. 1-22, 1999
- [2] Federal Railroad Administration (U.S. Department of Transportation), Dynamic Wheel Load Detector Extends Life of Concrete Railroad Ties, *TR news*, January-February, 1987
- [3] AEA Technology: Wheelchex® system.
<http://www.aeat.co.uk/rail/what/products/wheelchex.htm>
- [4] TagMaster, NedTrain, and Bass: Gotcha system-Wheel flat detection and axle load measurement. <http://www.tagmaster.com>
- [5] Straininstall: TrackAlert. <http://straininstall.com/rail.html>
- [6] Schenck measuring process systems: Multirail: WheelScan.
<http://www.schenck-process.de/catalogue/multirail.pdf>
- [7] Salient Systems Inc: (WILD) Wheel Impact Load Detector.
http://www.salientsystems.com/prod_wild.html
- [8] ØDS-CALTRONIC A/S: Wheel Monitoring System (WMS) detects impact defects and other tread irregularities.
<http://www.odegaard.dk/Caltronic/Default.htm>
- [9] Imagemap inc: wheelspec®. <http://www.imagemap.com/WheelSpec.html>
- [10] BHP institute of Railway Technology - Monash university: Wheel impact monitor (WIM). www.eng.monash.edu.au/railway
- [11] Jodon and inc: Microwave Train Wheel Defects Measurement System.
<http://www.jodon.com/wheel.html>



- [12] A. Bracciali and G. Cascini, A wheel flat detection device based on cepstrum analysis of rail acceleration measurements. *World congress on Railway Reserarch WCRR97*, Firenze, pp. 513-521, 1997
- [13] A. Bracciali and G. Cascini, Detection of corrugation and wheel flats of railway wheels using energy and cepstrum analysis of rail acceleration, *Journal of Rail and Rapid Transit*, 211, pp. 109-116, 1997
- [14] S. Lechowicz and C. Hunt, Monitoring and managing wheel condition and loading. *International Symposium on Transportation Recorders, May 3-5 1999*, 1999
- [15] Qibo Feng et al, A dynamic and quantitative method for measuring wheel flats and abrasion of trains. *15th World Conference on NondestructiveTesting (WCNDT)*, Rome, 2000
- [16] C.A. Papadopoulos, N.T. Niakas, and K. Diplari, Flatness identification in wheels of passing trains. *4th European Conference on Noise Control (Euronoise 2001)*, 2001
- [17] A. Bracciali, G. Lionetti, and M. Pieralli, Effective Wheel Flats Detection Through A Simple Device. *Proceedings of the 2002 Techrail Workshop*, Paris, 2002
- [18] D. Skarlatos, K. Karakasis, and A. Trochidis, Railway wheel fault diagnosis using a fuzzy-logic method, *Applied Acoustics*, 65, pp. 951-966, 2004
- [19] *TEKNIS ELECTRONICS: Wheel Condition Monitor (WCM)*.
<http://www.teknis.net>
- [20] B.P. Bogert, M.J.R. Healy, and J.W. Tukey, The quefrency analysis of time serie for echoes: Cepstrum, pseudo-autocovariance, cros-cepstrum and saphe cracking. *Proceedings of the symposium on time series analysis*, pp. 209-243, 1963
- [21] A. Bracciai and F. Piccioli, Attenuation of rail vibration: Analysis of experimental data. *Proceedings of the 5th World Congress on Railway Research*, Koln, 2001
- [22] D.J. Thompson, Experimental analysis of wavepropagation in railway tracks, *Journal of sound and Vibration*, 203(5), pp. 867-888, 1997



Author Index

Abril M.	235, 289	Cucala P.	805
Adamko N.	663	Cugnasca P. S.	319
Alberto A.	817		
Albrecht T.	207, 531, 541	Daamen W.	3
Allain L.	341	D'Ariano A.	531
Arias E.	817	Darius M. N.	167
Arisawa R.	189	de Almeida Jr. J. R.	139
Armstrong J.	479	Derito A.	923
Asuka M.	179	Distante A.	157
Auciello J.	859	Dremelj B.	713
		Drożdziel J.	849
Bae C. H.	757		
Bantin C.	115	El Koursi E. M.	147
Barber F.	235, 289		
Barnes M.	797	Fabbian F.	373
Belmonte F.	129	Fernandez A.	747, 805
Benet J.	817	Finlayson A.	735
Berhan D.	167	Fioravanti D.	859
Berkani K.	129	Friman B.	421
Bin N.	413	Fujimori A.	551
Böhringer F.	459	Fujinami K.	189
Bon P.	341, 351	Fukami K.	199
Bosso N.	829, 869	Fukuda M.	309
Bouch C.	479		
Boulanger J.-L.	55, 129, 351	Gannon M. J.	13
Bovy P. H. L.	3	Gauyacq D.	275
Braban C.	381	Geistler A.	459
		Georgescu M. P.	401
Cebrián D.	817	Ghantous-Mouawad M.	55
Chandesris M.	97	Ghoseiri K.	601
Chandramouli M.	33	Giannettoni M.	923
Chang K. Y.	895	Gimenes R. A. V.	139
Chang Sy.	895	Gnägi H. R.	329
Charon P.	381	Goel R.	693
Chen R. J. T.	895	Goodman C. J.	479, 735
Cheng K. H.	895	Gooßmann R.	621
Chou T.-Y.	33	Goverde R. M. P.	207, 631
Chung L. K.	33	Gugliotta A.	829, 869
Churchill G.	55		
Chymera M.	797	Hai G. C.	413
Cocci G.	449	Han M. S.	757
Cosulich G.	923	Hansen I. A.	565, 631
Cuartero F.	817	Hansen S.	45

Hartong M.....	693
Haruyama S.	683
Hatanaka T.....	199
Hei X.....	309
Hirai C.	551
Hiramoto T.	945
Honda S.	227
Hong S. H.	85
Hoogendoorn S. P.....	3
Hribar G.....	713
Huang B.....	33
Hwang J.-G.....	675, 703
Ingolotti L.....	235, 289
Iwata K.	309
Jadhav R.	391
Jaeger B.	255
Jank S.....	777
Jha M. K.	23
Jo H.-J.....	675, 703
Kaas A. H.	45, 643
Kaneko S.....	299
Kasper J.	105
Kavička A.....	663
Kim Y. K.	757
Klier F.....	777
Klima V.	663
Ko H.	767, 787
Kojima T.....	245, 955
Komaya K.....	179
Kondou S.....	551
Kotake H.....	683
Kufver B.	839
Kwon S. Y.	757
Labropoulos K. C.....	935
Lahlou O.....	341
Landex A.	45, 65, 521, 643
Lee H. U.	85
Lee J. H.....	85, 675
Lemmer K.....	255
Lindfeldt O.	511
Lindqvist L.	391
Liu B.....	965

Liu Y.....	965
Loeve L.....	501
Lova A.....	235, 289
Lukaszewicz P.	723
Maeda T.....	469
Malavasi G.....	905
Malvezzi M.....	449
Mariano G.....	351
Matsuda K.	767
Matsumoto A.	955
Matsuzawa T.	683
Mazzeo P. L.....	157
Meier-Hirmer C.	217
Middelkoop A. D.....	501
Min Q. K.....	413
Miwa M.	885
Miyatake M.	767, 787
Mizuma T.	945
Mochizuki H.....	299, 309, 945
Moropoulou A.	935
Moundoulas P.....	935
Muñoz-Riesco J.	805
Myojo S.	653
Nagasaki Y.	179
Nakagawa M.....	683
Nakagawa T.....	189
Nakamura H.....	299, 309, 945
Nielsen O. A.	521
Nitti M.	157
Noguchi Y.	299
Nogueira T. R.	139
Noury P.....	431
Obrenovic M.....	255
Oh S. M.	85
Ono M.....	469
Park B. H.	85
Park H. J.	757
Persson R.....	839
Pilo E.	747, 805
Polzin T.	621
Preston J.	479
Pugi L.	859

Quattrini A.	905	Takagi R.	479
Rachel F. M.	319	Takahashi S.	299, 309, 945
Radtke A.	579	Tanabe H.	227
Raicu R.	75	Tao T.	413
Renfrew A. C.	797	Tarasconi A.	449
Riboulet G.	217	Tashiro Y.	551
Ricci S.	905	Taylor M. A. P.	75
Rinchi M.	859	Terada T.	199
Rindi A.	859	Thomas P.	363
Rodriguez J.	591	Tomii N.	551
Rojo T.	817	Tordai L.	147
Rouco L.	747	Tormos P.	235, 289
Roussignol M.	217	Tsuchiya R.	189, 265
		Tsunashima H.	245, 955
Sakai M.	299	Ukai M.	265
Sakowitz C.	489	Valliyappan N.	167
Salido M. A.	235, 289	van Luipen J.	207
Salling K. B.	65	Vinolas J.	973
Samanta S.	23		
Sanchez E.	973	Watanabe K.	469
Sato K.	309	Weeda V. A.	207, 611
Savio S.	923	Wendler E.	489
Schittenhelm B.	643	Weston P. F.	479
Schneider-Tilli J.	643	White R. D.	735
Scholz S.	541	Wiggenraad P. B. L.	611
Schön W.	55, 129	Wijesekera D.	693
Seco M.	973	Wolbers M. J.	915
Sekavčnik M.	713	Wu Q.	33
Seki K.	265		
Selig T.	105	Xie J.	777
Senée A.	217		
Shimizu H.	227	Yamamoto H.	199
Shimura A.	683	Yamauchi K.	189
Shiozawa T. Y.	245	Yasui Y.	441
Solomon G.	167	Yasura K.	227
Somà A.	829, 869	Yoon Y.-K.	703
Sone S.	479	Yoshimura A.	885
Sourget F.	217	Yuan J.	631
Sowiński B.	849		
Stahel N.	329	Zhao H.	965
Stella E.	157	Zynovchenko A.	777
Sugiyama Y.	189		
Suzuki S.	265		



WITPRESS

Intelligent Road Design

*M. K. JHA, Morgan State University, USA,
P. M. SCHONFELD, University of
Maryland, USA, J-C. JONG, Sinotech
Engineering Consultants, Inc., Taiwan,
R.O.C., E. KIM, The Korea Transport
Institute, South Korea*

Traditional textbooks on roadway designs focus on fundamental alignment elements and design criteria. When applying these in practice, engineers must try various design combinations, check whether the resulting alignments satisfy requirements, and evaluate their relative effectiveness.

Introducing a systematic and efficient approach to optimize alignments, this essentially practical text emphasizes the use of artificial intelligence (AI) and Geographic Information Systems (GIS) in extensively automated highway design. Based on a series of research projects, it provides a thorough introduction to the mathematical models and solution algorithms for optimizing highway alignments, including horizontal, vertical, and three-dimensional alignments.

The text is ideally suited to senior undergraduate or graduate students majoring in civil engineering or transportation management. Practicing highway design and transportation engineers will also find it of interest.

Series: Advances in Transport Vol 19

ISBN: 1-84564-003-9 2005 448pp

£147.00/US\$264.00/€220.50

Fuel Efficient Car Technology

M.L. POULTON, Automotive Consultant

Presents measures designed to reduce fuel consumption in passenger cars.

ISBN: 1-85312-447-8 1997 160pp

£59.00/US\$91.00/€88.50

Urban Transport XII

Urban Transport and the Environment in the 21st Century

*Edited by: C. A. BREBBIA, Wessex
Institute of Technology, UK,
V. DOLEZEL, Tu Pardubice, Czech
Republic*

Transportation in cities, with its related environmental and social concerns continues to be a topic of the utmost priority for urban authorities and central governments around the world. This is reflected in the proceedings of the Twelfth International Conference on Urban Transport and the Environment in the 21st Century, stressing the continuous steady growth and research into the urban transport systems control aspects, information and simulation systems. Papers cover topics such as: Transport Logistics and Operations Research; Transport Modelling and Simulation; Intelligent Transport Systems; Urban Transport Planning and Management; Road and Parking Pricing; Public Transport Systems; Environmental and Ecological Considerations; Transport Sustainability; Infrastructure and Maintenance; Information Systems and GPS Applications; Transport Security and Safety; Transport Technology; Energy and Transport Fuels; Land Use and Transport Integration.

*WIT Transactions on The Built
Environment, Volume 89*

ISBN: 1-84564-179-5 2006 apx 600pp

apx £215.00/US\$385.00/€322.50

Find us at

<http://www.witpress.com>

**Save 10% when you order from our encrypted
ordering service on the web using your credit card.**



WITPRESS

Advances in City Transport

Case Studies

*Edited by: S. BASBAS, Aristotle University
of Thessaloniki, Greece*

Highlighting the highly topical subject of transport and the environment and the closely related field of town planning, this book contains chapters concerning developments in the transportation systems of various cities all over the world. These include Singapore, Sao Paulo, Santiago, Bilbao, Eindhoven, Adelaide, Bangalore and Thessaloniki.

The studies featured will be of interest to postgraduate researchers in transport and the environment, engineers and planners working within transport and environment ministries and local authorities, and consultants.

Series: Advances in Transport Vol 17

**ISBN: 1-85312-799-X 2005 apx 224pp
apx £85.00/US\$129.00/€127.50**

How to Make Two-Lane Rural Roads Safer

Scientific Background and Guide for Practical Application

*R. LAMM, University of Karlsruhe,
Germany, A. BECK, University of
Karlsruhe, Germany, T. RUSCHER,
University of Karlsruhe, Germany,
T. MAILAENDER, Mailaender Ingenieur
Consult GmbH, Germany, S. CAFISO,
University of Catania, Italy, G. LA CAVA,
University of Catania, Italy*

In most countries two-lane rural roads make up about 90 percent of rural networks and account for about 60 percent or more of

highway fatalities worldwide - 500,000 people per year. Based on new research and the demands of many design professionals this book provides an understandable scientific framework for the application of quantitative safety evaluation processes on two-lane rural roads.

The methodology described will support the achievement of quantified measures of 1) design consistency, 2) operating speed consistency, and 3) driving dynamic consistency. All three criteria are evaluated in three ranges described as "good", "fair" and "poor". It has been proved that the results of these criteria coincide with the actual accident situation prevailing on two-lane rural roads. By using the "good" ranges sound alignments in plan and profile, which match the expected driving behaviour of motorists, can be achieved.

The safety criteria are then combined into an overall safety module for a simplified general overview of the safety evaluation process. The authors also encourage the coordination of safety concerns with important economic, environmental and aesthetic considerations. A CD-ROM with practical applications accompanies the text.

This book will be an invaluable aid to educators, students, consultants, highway engineers and administrators, as well as scientists in the fields of highway design and traffic safety engineering.

ISBN: 1-84564-156-6 2006

apx 104pp+CD-ROM

apx £59.00/US\$94.00/€88.50

WITPress

**Ashurst Lodge, Ashurst, Southampton,
SO40 7AA, UK.**

Tel: 44 (0) 238 029 3223

Fax: 44 (0) 238 029 2853

E-Mail: marketing@witpress.com





WITPRESS

Urban Transport XI

Urban Transport and the Environment in the 21st Century

*Edited by: C. A. BREBBIA, Wessex Institute of Technology, UK,
L. C. WADHWA, James Cook University, Australia*

The continuing need for better urban transport systems and a healthier environment has led to an increased level of research around the world. This is reflected in Urban Transport XI, which features the proceedings of the latest conference in this well-established series. The subjects covered are of primary importance for analysing the complex interaction of the urban transport environment and for establishing action strategies for transport and traffic problems. Over 85 papers are included and these highlight topics within the following areas: Urban Transport Systems; Public Transport Systems; Infrastructure and Maintenance; Safety and Security; Transport Sustainability; Accessibility and Mobility; Environmental Impacts; Air and Noise Pollution; Energy and Fuel; Integrated Land Use and Transport; Travel Demand Management; Traffic Control and Integration; Advanced Transport Systems; Simulation; Economic and Social Impacts; and Cost and Investment Analysis.

WIT Transactions on The Built Environment, Volume 77

ISBN: 1-84564-008-X 2005 928pp
£297.00/US\$475.00/€445.50

We are now able to supply you with details of new WIT Press titles via

E-Mail. To subscribe to this free service, or for information on any of our titles, please contact the Marketing Department, WIT Press, Ashurst Lodge, Ashurst, Southampton, SO40 7AA, UK

Tel: +44 (0) 238 029 3223

Fax: +44 (0) 238 029 2853

E-mail: marketing@witpress.com

Innovations in Freight Transport

Editors: E. TANIGUCHI, Kyoto University, Japan and R.G. THOMPSON, University of Melbourne, Australia

Highlighting new ideas and best practice, this book examines innovations in modern freight transport systems.

Partial Contents: Intelligent Transport Systems; Vehicle Routing and Scheduling; Logistics Terminals; Intermodal Freight Transport; Underground Freight Transport Systems; E-Commerce and the Consequences for Freight Transport; Future Perspectives.

Series: Advances in Transport, Vol 11

ISBN: 1-85312-894-5 2002 216pp
£76.00/US\$118.00/€114.00

Urban Transport X

Urban Transport and the Environment in the 21st Century

*Editors: C. A. BREBBIA, Wessex Institute of Technology, UK and
L.C. WADHWA, James Cook University, Australia*

The study of urban transport is complex and involves many disciplines as shown by the papers in this book. These come from the Tenth International Urban Transport Conference and cover topics such as energy systems, transport systems and integration, and environmental impact.

WIT Transactions on The Built Environment, Volume 75

ISBN: 1-85312-716-7 2004 936pp
£309.00/US\$494.00/€463.50

All prices correct at time of going to press but subject to change.

WIT Press books are available through your bookseller or direct from the publisher.