Jerzy Mikulski (Ed.)

Communications in Computer and Information Science 104

Transport Systems Telematics

10th Conference, TST 2010
Katowice – Ustroń, Poland, October 2010
Selected Papers

Springer

Communications
in Computer and Information Science 104

Jerzy Mikulski (Ed.)

Transport Systems Telematics

10th Conference, TST 2010
Katowice – Ustroń, Poland, October 20-23, 2010
Selected Papers

Springer

Volume Editor

Jerzy Mikulski
Politechnika Śląska
ul. Krasińskiego 8/201
40-019 Katowice, Poland
E-mail: jerzy.mikulski@polsl.pl

Library of Congress Control Number: 2010936519

CR Subject Classification (1998): C.2, H.4, H.3, H.2, H.5, D.2

ISSN 1865-0929
ISBN-10 3-642-16471-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-16471-2 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

The idea of telematics appeared more than a decade ago and it is possible to define it, in a general and simple way, as a communication system for collecting, processing and distributing information.

The transport services market is definitely the most important area for telematic applications. Transport-telematics issues constitute a field of knowledge of transport that integrates information technology and telecommunications in applications for managing and controlling traffic in transport systems, stimulating technical and organizational activities that ensure improved effectiveness and safe operation of such systems. Integrated and cooperating telematic applications constitute intelligent transport systems. The basis of such systems is to efficiently collect and process information and to manage its flow within the system. This enables supplying information from almost all areas of transport activities in real time.

Intelligent transport—supported by a number of integrated telecommunications, IT measurement and control engineering solutions, and by appropriate tools and software—comprises telematic applications. They have an extensive range of use in many areas of transport, allowing the integration of the means and types of transport, including its infrastructure, business organization and management processes.

This monograph is a collection of selected papers presented at the jubilee transport telematics conference, TST 2010, and is the result of the work of many scientists associated with this area of knowledge and who had spent years with the conference. The conference was organized for the tenth time, the only scientific conference with so much experience in this field of science in Poland. It should be noted that since inception the conference has had the status of an international conference.

I would like, on behalf of the Scientific Programme Committee, to thank the authors for their sizeable contribution to the knowledge in the field of transport telematics and in the creation of this book.

October 2010 Jerzy Mikulski

Organization

Organizers

- Department of Railway Engineering, Faculty of Transport,
  Silesian University of Technology, Katowice, Poland
- Polish Academy of Sciences, Transport Committee, Poland
- Polish Association of Transport Telematics, Poland

Co-organizers

- Faculty of Transport, Warsaw University of Technology, Poland
- Faculty of Transport and Electrical Engineering,
  Radom University of Technology, Poland
- Faculty of Economics, University of Economics in Katowice, Poland
- Faculty of Social and Technical Sciences,
  Silesian High School of Management in Katowice, Poland
- Regional Center of Road Traffic, Katowice, Poland
- Motor Transport Institute in Warsaw, Poland
- Railway Institute, Warsaw, Poland
- Civil Aviation Personnel Education Center of Central and Eastern Europe,
  Katowice, Poland

Organizing Committee

Chair: J. Młyńczak
Members:

- A. Białoń,
- P. Gorczyca,
- K. Kędziora,
- J. Łukasik,
- S. Surma

Secretary: R. Skowrońska

Conference Took Place under the Honorable Patronage of:

- Minister of Infrastructure
- Rector of Silesian University of Technology
- President of Polish Air Navigation Services Agency

- ERTICO - ITS Europe
- Association of Polish Electrical Engineers SEP
- IEEE Poland Section

Scientific Program Committee

- J. Mikulski - chairman, Silesian University of Technology, Poland
  Polish Association of Transport Telematics, Poland
- A. Lewiński - vice-chairman, Technical University of Radom, Poland
- W. Wawrzyński - vice-chairman, Warsaw University of Technology, Poland

- R. Bańczyk, Regional Center of Road Traffic, Poland
- M. Bregulla, University of Applied Sciences, Ingolstadt, Germany
- K. Chwesiuk, Maritime University of Szczecin, Poland
- J. Dyduch, Technical University of Radom, Poland
  Polish Academy of Sciences, Poland
- A. Fellner, Silesian University of Technology, Poland
- W. Filipowicz, Gdynia Maritime University, Poland
- M. Franeková, University of Zilina, Republic of Slovakia
- W. Galor, Maritime University of Szczecin, Poland
- S. Gucma, Maritime University of Szczecin, Poland
- M. Holec, Gdynia Maritime University, Poland
- A. Janota, University of Zilina, Republic of Slovakia
- J. Januszewski, Gdynia Maritime University, Poland
- A. Kalašová, University of Zilina, Republic of Slovakia
- D. Kevicky, University of Zilina, Republic of Slovakia
- J. Klamka, Polish Academy of Sciences, Poland
- B. Kos, Academy of Economics in Katowice, Poland
- O. Krettek, RWTH Aachen, Germany
- R. Krystek, Gdansk University of Technology, Poland
- J. Lewitowicz, Air Force Institute of Technology, Warsaw
- M. Luft, Technical University of Radom, Poland
- B. Łazarz, Silesian University of Technology, Poland
- Z. Łukasik, Technical University of Radom, Poland
- H. Madej, Silesian University of Technology, Poland
- J. Michalík, University of Zilina, Republic of Slovakia
- R. Michalski, Silesian School of Management, Katowice, Poland
- I. Mitraszewska, Motor Transport Institute, Warsaw, Poland
- J. Modelski, IEEE Region 8
- G. Nowacki, Motor Transport Institute, Warsaw, Poland
- S. Oszczak, University of Warmia and Mazury in Olsztyn, Poland
- M. Pawlik, Polish Railway Lines, Poland
- Z. Pietrzykowski, Maritime University of Szczecin, Poland

- B. Pochopień, Silesian University of Technology, Poland
- K. Rástočný, University of Zilina, Republic of Slovakia
- M. Sitarz, Silesian University of Technology, Poland
- J. Spalek, University of Zilina, Republic of Slovakia
- R. Srp, ITS&S, Czech Republic
- Z. Stotsko, Lviv Polytechnic National University, Ukraine
- W. Suchorzewski, Warsaw University of Technology, Poland
- M. Svítek, ITS&S, Czech Republic
- A. Szeląg, Warsaw University of Technology, Poland
- E. Szychta, Technical University of Radom, Poland
- G. Tarnai, Technical University of Budapest, Hungary
- R. Tomanek, University of Economics in Katowice, Poland
- Z. Toš, University of Zagreb, Croatia
- R. Wawruch, Gdynia Maritime University, Poland
- A. Weintrit, Gdynia Maritime University, Poland
- B. Wiśniewski, Maritime University of Szczecin, Poland
- A. Wojciechowski, Motor Transport Institute, Poland
- K. Wydro, University College of Technology and Business, Poland
- E. Załoga, Maritime University of Szczecin, Poland

Referees

Andrzej Fellner
Mária Franeková
Aleš Janota
Alica Kalašová
Barbara Kos
Stanisław Krawiec
Otmar Krettek
Andrzej Lewiński
Jerzy Mikulski
Jakub Młyńczak
Gabriel Nowacki
Zbigniew Pietrzykowski
Adam Weintrit
Kornel Wydro

Table of Contents

Telematic Approach to e-Navigation Architecture

Adam Weintrit

Gdynia Maritime University, The Faculty of Navigation, Al. Jana Pawla II 3, 81-345 Gdynia, Poland
weintrit@am.gdynia.pl

Abstract. e-Navigation is an IMO (International Maritime Organization) initiative defined as “the harmonised collection, integration, exchange, presentation and analysis of maritime information onboard and ashore by electronic means to enhance berth to berth navigation and related services, for safety and security at sea and protection of the marine environment”. In the paper the Author tries to discuss a telematic approach to e-Navigation architecture and stress that for enhancement of safety, security and environmental protection, the decision makers must be supported not only by the technology, but also in a significant way with effective procedures and training.

Keywords: Maritime Transportation, Navigation, Telematics, e-Navigation.

1 Introduction

The advantage of the latest technical development in the field of automation, electronics, telecommunications, informatics, telematics, geomatics and global position fixing techniques, achievements in data storing, processing, analysing, transferring and visualisation should be taken into account and applied to the maritime technology [7]. In the paper the Author tries to discuss the main tasks of the maritime community for the near future in the field of e-Navigation.

It is now appropriate to develop a broad strategic vision for incorporating the use of new technologies in a structured way and ensuring that their use is compliant with the various electronic navigational and communication technologies and services that are already available [5]. The aim is to develop an overarching accurate, safe, secure and cost-effective system with a potential to provide global coverage for vessels of all sizes [2]. The implementation of this new strategic vision might require modifications to working methods and navigational tools, such as inner ship’s computer net, charts, bridge display equipment, electronic aids to navigation, communications and shore infrastructure. At this stage, it is difficult to be precise about the full extent of the changes that might be necessary to fully deliver this vision. However, changes to a number of regulatory instruments might be needed, including the appropriate chapters in the SOLAS Convention. This proposal is not in any way intended to conflict with the clear principle, as confirmed in SOLAS, of the master’s authority for the operational safety of the vessel, and in UNCLOS, of freedom of navigation rights.

Fig. 1 displays the conceptual process of e-Navigation. The legal requirements ”influence” the whole process. Training and operational procedures are affected by: the

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 1–10, 2010.

e-Navigation process, the users' needs, the operational functions to be carried out, and the technical facilities. The user needs and especially the operational functions are influenced by the safety management of the company, the culture, and depend on the ship type and the equipment configuration, and issues of quality and reliability must be addressed throughout the process in terms of data systems and training.

The iterative process and the dependency of the technical facilities are symbolized with the "circle arrows".

It is also certain that the safe and efficient transport will continue to rely on good decisions being made on an increasingly constant and reliable basis. Some decisions may be made with increased dependence on technology, but at some level we will always rely on good human decisions being made and therefore every effort needs to be made to apply an understanding of the Human Element at all stages, of design, development, implementation and operation of e-Navigation. The Nautical Institute [6] as the leading international body for maritime professionals will continue to use the resources of its members, branches, officers and staff to promote the effective application of the Human Element for e-Navigation and other industry developments, and invites all maritime professionals to join in this critical effort.

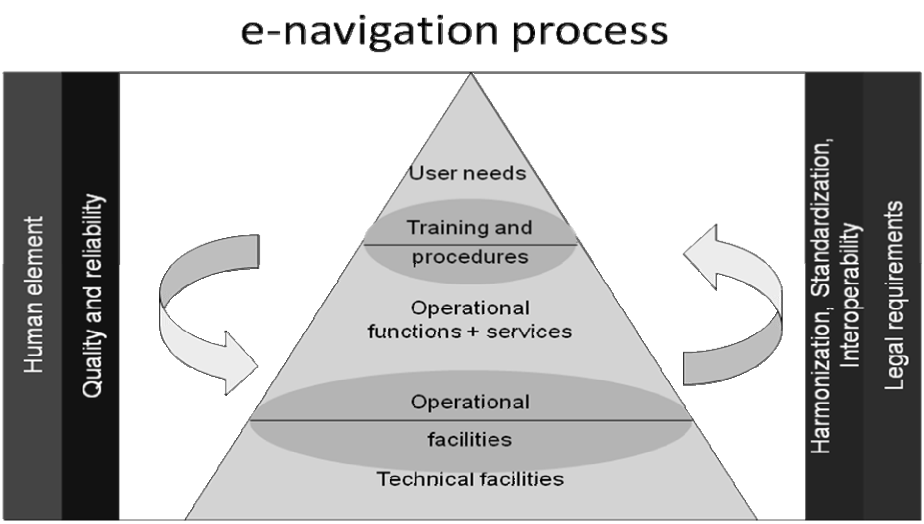


Fig. 1. e-Navigation: structure of the conceptual process [4]

2 Introduction to e-Navigation Architecture

The work of implementing the e-Navigation strategy should begin by the outline of architecture [3]. To become workable, the current definition of e-Navigation needs to be broken down into detail. The work is complex and will require input from a number of experts. For reasons of efficiency and effectiveness the work needs to be conducted in a systematic and consistent way. There is therefore a need to develop an architecture, a framework, within which the definition of e-Navigation as presented may be further refined. The detailed definition will address not only the concept and functions of e-Navigation which are non-technical, but also its technical components, as well as the complex relations between these [1].

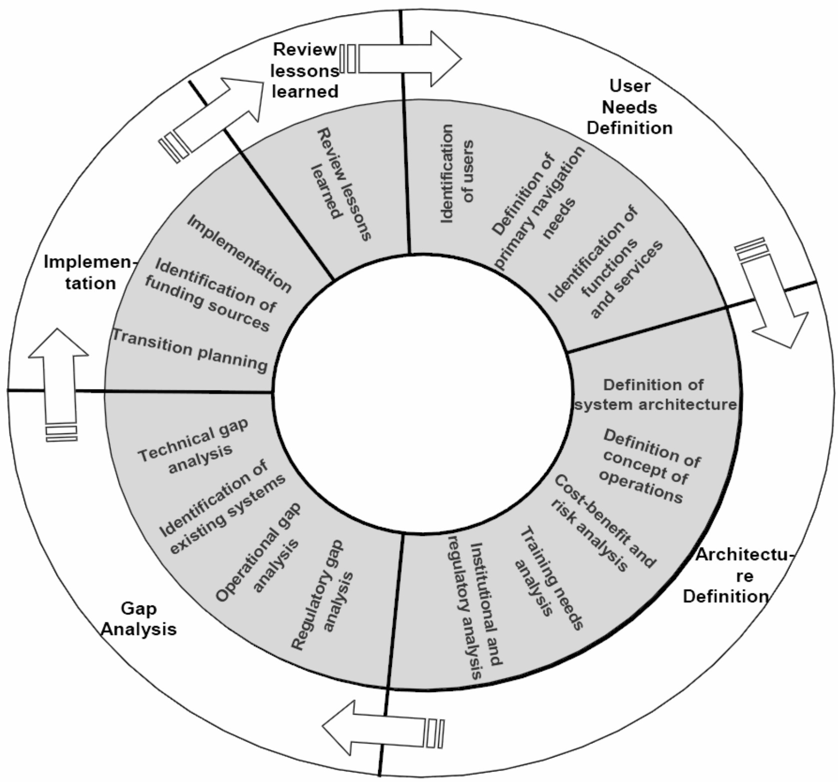


Fig. 2. Potential components of an e-Navigation implementation plan [6]

The architecture serves two purposes [4]:

1. The first is to ensure a common understanding and interpretation of the e-Navigation concept. Although documents [1], [3], and [4] provide an overall definition of e-Navigation, a more detailed definition is required. This definition must be sustainable, meaning that it as far as possible must be stable over time and independent of technology, technical components and systems (which may change over time). The definition should further clarify:
   - The context of e-Navigation. That is the main aspects, scope and environment of e-Navigation and the responsibilities (i.e. roles) to be fulfilled by stakeholders.
   - The logical description of e-Navigation by means of functions that enables stakeholders to meet their responsibilities; and the information that has to be exchanged between these functions. Responsibilities/roles, functions and

information flows should be combined into processes related to specific situations and stages of the ship's voyage.
2. The second purpose is to support the implementation of the e-Navigation concept. The architecture should provide implementation requirements with references to the logical specifications, e.g. specify how to carry out the information flows. The current technologies, standards and solutions to be used should be defined.

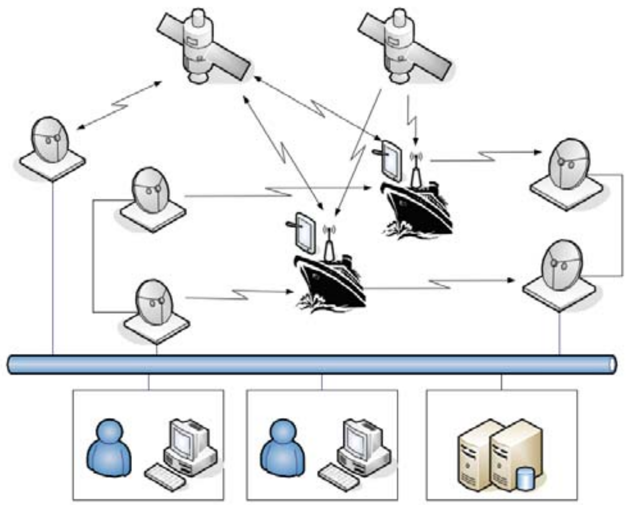


Fig. 3. General telematic approach to e-Navigation [9]

The architecture defines the scope of e-Navigation, its boundaries to the external environment as well as the responsibilities/roles to be met by stakeholders. It describes the information exchange between various functions. Finally it describes the processes in which responsibilities, functions and information flows between these are combined and related to a specific situation, e.g. stages of the ship's voyage.

The technical solutions will follow the conceptual specifications. It may however take time to establish all the implementation requirements because standards and solutions may not yet be available. Therefore at the outset the implementation specifications may not be as complete as the conceptual specifications. These will have to be updated as new technologies, standards and solutions emerge.

The initial technical e-Navigation architecture work is based on the understanding of IMO's e-Navigation concept in general, and telematic approach to e-Navigation architecture specifically.

A sufficiently advanced proper infrastructure needs to be present. This infrastructure would consist of supporting devices, which e.g. allow the application software to be executed (i.e. computers, computer peripherals, operating systems, local area network components, etc.). The infrastructure itself becomes more and more sophisticated and is subject to its own requirements, such as reduction of energy consumption and environmental friendliness.

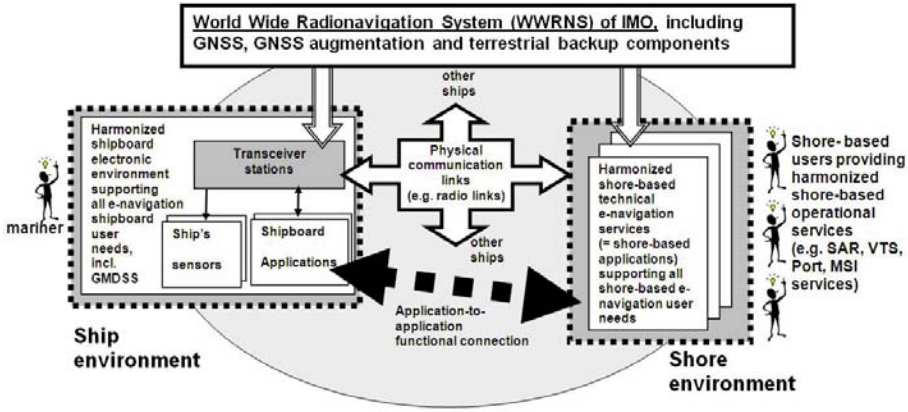


Fig. 4. Conceptual e-Navigation architecture [4]

One important principle is the orientation towards the information flow instead of technology development. That principle is represented by Fig. 4 below, which shows the shipboard environment, the physical link(s) and the shore-based environment.

Fig. 4 is a representation of the e-Navigation environment. On the left there is a single "ship environment". From an e-Navigation perspective, the relevant devices within the ship environment are the transceiver station, the sensors and applications connected to the transceiver station, the Integrated Navigation System (INS) and the Integrated Bridge System (IBS). The GMDSS function has been included in the ship environment, with a direct link to the transceiver station and a link to the World Wide Radio Navigation System (WWRNS).

The transceiver station is shown as a single station although in reality there may be several transceiver stations. It interfaces via links with the appropriate technical e-Navigation services ashore.

As regards shore-based services there are technical services responsible for interfacing with the ship services. Operators ashore, such as VTS, port, pilot station and SAR operators, or lock operators, perform their tasks in co-operation with shipboard applications. From the shore perspective, it is the functional links between the shore-based user applications and the shipboard applications that are most relevant, such as the provision of MSI.

A similar setup of interactions applies for ship-to-ship and shore-to-shore applications, and would include, for example, real-time and near real-time meteorological and oceanographic services. Pilots also perform the same interactions as mariners before and when they are on board ships in pilotage waters. The above functional links are equally important to ensure efficient ship to ship, shore to ship and ship to shore communications. Fig. 4 also shows dependency on the WWRNS.

The bold arrow in Fig. 4 represents the data exchange between the shore-based applications and the "ship's environment", and vice versa. Physical links between (fixed) shore and (mobile) shipboard equipment each employ one or more appropriate methods such as radio or light signals.

3 The Concept of a Common Data Structure

IMO, IHO (International Hydrographic Organization) and IALA (International Association of Lighthouse Authorities) have raised the issue that some form of common data structure, representing the maritime domain (and including both ship and shore aspects), will be essential for e-Navigation.

In January 2010, the IHO introduced a new data model to be known as S-100 - the Universal Hydrographic Data Model (UHDM). This international standard has been developed by the IHO over the last nine years in consultation with a wide range of stakeholders, including key ECDIS and navigation equipment manufactures. The purpose of S-100 is to provide a framework architecture for a contemporary standard for the exchange of hydrographic and related maritime data. S-100 is based on the ISO 19100 series of geographic standards and is fully compatible and interoperable with those standards. As the UHDM is aligned with ISO 19100 it will enable the exchange of hydrographic and other maritime data and information together with the geospatial data from other domains. The use of data standards enables interoperability between geospatial data sets from different domains and could therefore be appropriate for many of the datasets envisaged for data exchange in e-Navigation.

S-100 is not limited to the hydrographic data or hydrographic applications. It has been developed specifically to enable the widest possible range of users to use the hydrographic data in conjunction with the data from other maritime and marine domains. Like traditional applications such as nautical charts and publications, applications based on S-100 already under development by non-IHO stakeholder groups include the sea ice forecast reporting, recording UNCLOS boundaries, and marine information overlays. These are applications that obviously encompass various hydrographic, meteorological and oceanographic parameters that go well beyond the traditional navigation and hydrographic products provided by HO's. S-100 is intended to be a fundamental enabler for hydrographic input to Marine Spatial Data Infrastructures (MSDI) as well as for other developing marine information infrastructures such as e-Navigation.

IALA is currently developing a proposed Universal Maritime Data Model (UMDM) for e-Navigation to meet requirements arising from the future implementation of e-Navigation. It is therefore important to harmonize efforts in data modelling, with the aim of creating and maintaining a robust and extendable maritime data structure.

The common maritime information and data structure will require some form of overarching coordination to ensure the ongoing management and maintenance of the structure. There may be several management roles to be performed by such a coordinating body, (for example, the maintenance of a register). This management role may be able to be shared between relevant organisations.

The common data structure should contain data models like IALA's UMDM, IHO's UHDM and data models of other international stakeholders. The structure is a highly important element by which e-Navigation can modernize the operational environment of the maritime industry. Reporting should be standardized and in a format that supports the effective use ashore, such as a global voluntary single window network. Construction of the UMDM will be a collaborative effort among many parties involved in the maritime environments. By having each party bringing its particular expertise, the UMDM will become the accepted standard model.

Implementation of IMO's e-Navigation strategy leads to a larger variety and higher volume of information and increased information exchange due to globalization. Consequently there is a need to handle the information more effectively in a standardized way. The first step towards a common data structure is to define the meaning of each and every item in the data structure and the relationships among the items. This is done so that implementers of the data structure have common understanding of items. The means to do this is with a data model. At this stage, the data model, like the system architecture needs only to be described in the most general of terms. An example of how a UMDM could be implemented has been provided by the IHO with its S-100 model.

4 e-Navigation Architecture Structure

NAV 53/13 [3] states that "e-Navigation can be described in terms of its component elements, as a process model identifying inputs and outputs, or in relation to the different stages of a ship's voyage". NAV 53/13 also presents three diagrams that take an approach to the design of architecture [3]:

- An integrated bridge system diagram illustrating stakeholders, functions/processes, information, technologies and systems
- A diagram identifying inputs and outputs to e-Navigation, functions/processes and benefits
- A diagram from the IALA's E-NAV Committee on the safety of navigation identifying functions/processes, information, technologies and systems [1].

The IMO correspondence group on e-Navigation recommends that the elements depicted in the diagrams be used as inputs to the design of the architecture. The different types of components (functions, information, processes, etc.) must however be organised into viewpoints addressing different aspects of e-Navigations such as the functions required, the information to be exchanged and processes describing the activities in specific situations or at different stages of the ship's voyage. Standards should be stated as the backbone and one of the major pillars of functionalities and interoperability of all the systems and information.

The architecture is organised into 3 abstraction levels as illustrated in Fig. 5. Levels 1 and 2 are non-technical. They are prerequisites for a common understanding of the e-Navigation concept. Level 3 defines the technical realisation and supports the implementation of the e-Navigation concept.

The layered model and the different viewpoints depicted in Fig. 5 will give a well defined structure to the specification of e-Navigation and also promote the establishment of clear definitions, a common understanding of the concept and well defined terminologies. This will support the requirements to the architecture.

The technical implementations of the solutions defined at Level 2 are specified by means of system components and communication solutions.

The architecture and the e-Navigation concept should be based on a consolidation of user needs across the entire range of users, taking into account all possible economies of scale. The architecture should include hardware, data, information, communications and software needed to meet the user needs. The structured and layered approach to architecture as illustrated in Fig. 5 will support the architecture development process.

User requirements can be related to roles (i.e. responsibilities) reflecting the entire range of users. The viewpoints will support the capturing of user requirements and the interaction with them in the development of requirements. Existing user requirements can be entered into the respective viewpoints and be the starting points for further discussions with users and maritime experts. The process viewpoint will support the verification of the other components - the roles, the functions and the information and lead to a consistent result [8].

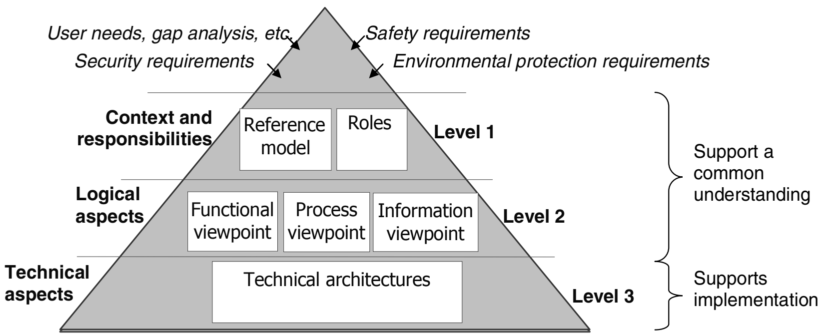


Fig. 5. e-Navigation architecture structure and components [4]

4.1 Architecture Components Should Be Made Modular and Scalable

The architecture should be based on a modular and scalable concept. The hardware and software should be based on open architectures to allow scalability of functions according to the needs of different users to cater for a continued development and enhancement.

The roles (described below) will arrange for a flexible organisation of e-Navigation solutions. Responsibilities, and the associated functions and processes, can be adapted to local stakeholders in a flexible way, or the responsibilities may be handled on a regional basis. Thus the architecture arranges for local, national and regional solutions.

The technical solutions must ensure that the system components are modular and scalable (e.g. adaptable to traffic density, etc.). The logical parts of the architecture will however be a good starting point for the technical aspects by providing a structured specification of the solutions. The information viewpoint will for example provide specifications of the information content to be exchanged through standardised and open interfaces.

4.2 Architecture Should Be Adaptable to New Requirements

The architecture should include the hardware, data, information, communications technology and software needed to meet the user needs. The system architecture should be based on a modular and scalable concept. The system hardware and software should be based on open architectures to allow scalability of functions according to the needs of different users and to cater for continued development and enhancement. The requirement is met by making a separation between the contextual and logical parts and the technical part of the architecture. The non-technical parts will ensure consistent and common understanding of the solutions, and the technical specifications will provide implementation guidelines and requirements by means of the available technologies, standards and solutions. The technical aspects may be updated as new technologies, standards and solutions emerge.

4.3 Architecture Should Support Training Needs Assessment

Training needs analysis should be performed based on the system architecture and the operational concept resulting in a training specification. The process viewpoint specifying processes related to specific situations and stages of the voyage can be used as a starting point for training scenarios. In addition, the technical aspects will identify technologies and systems that should be addressed in the training of the users.

The already defined roles support the addressing of specific responsibilities, and this can also be reflected by the processes. Thus, it should be possible to customise the training scenarios to the individual stakeholders depending on their roles.

4.4 Architecture Should Support Institutional and Regulatory Analysis

Institutional and regulatory requirements' analysis should be undertaken, based on the system architecture and operational concepts.

The roles and the viewpoints will be a tool in the analysis of responsibilities and processes (including functions and information exchange), and the technical aspects can be used to define implementation plans and requirements.

It is also likely that the work with regulations related to the implementation of e-Navigation may benefit from the architecture. The terminology defined by the architecture should be used. This will ensure consistency with respect to terminology across regulations for example with respect to stakeholders and responsibilities and the architecture will support the interpretation of the regulations.

5 Conclusions

e-Navigation is a broad concept that is aimed at enhancing navigation safety, security and the protection of the marine environment through a harmonised collection, integration, exchange, presentation and analysis of maritime information onboard and ashore by electronic means. A telematic approach to e-Navigation is very important.

It is envisioned that e-Navigation will be a 'living' concept that will evolve and adapt over a long time scale to support this objective. During this time the information will change, technologies will change, political and commercial objectives will change, and tasks will change. However it is unlikely that the need for safe and efficient seaborne transport will change significantly.

The overall conceptual, functional and technical architecture will need to be developed and maintained, particularly in terms of process description, data structures, information systems, communications technology and regulations. The architecture should include the hardware, data, information, communications technology and software needed to meet the user needs. The system architecture should be based on a modular and scalable concept. The system hardware and software should be based on open architectures to allow scalability of functions according to the needs of different users and to cater for continued development and enhancement. The architecture should contribute to the common understanding and interpretation of the e-Navigation concept. The current definition needs to be broken down into a detailed definition/description of e-Navigation. The description should be made sustainable, i.e. stable over time. It further needs to have the capacity to integrate developments in technical components and systems.

References

1. IALA. The IALA definition and vision for e-Navigation. E-NAV2-output 11 (March 2007)
2. Mitropoulos, E.: e-Navigation: a global resource. Seaways. The International Journal of the Nautical Institute (March 2007)
3. NAV 53/13. Development of an e-Navigation Strategy. Report of the Correspondence Group on e-Navigation, submitted by the United Kingdom. Sub-Committee on Safety of Navigation, International Maritime Organization, London (April 20, 2007)
4. NAV 56/8/... Development of an e-Navigation Strategy Implementation. Report of Correspondence Group, submitted by Norway. Sub-Committee on Safety of Navigation, International Maritime Organization, London (April 23, 2010)
5. Patraiko, D.: The Development of e-Navigation. In: Weintrit, A. (ed.) Advances in Marine Navigation and Safety of Sea Transportation, ch. 10: e-Navigation. TransNav 2007 Monograph. Gdynia Maritime University and the Nautical Institute, Gdynia (2007)
6. Patraiko, D., Wake, P., Weintrit, A.: e-Navigation and the Human Element. In: Weintrit, A. (ed.) Monograph. Marine Navigation and Safety of Sea Transportation. A Balkema Book, CRC Press, Taylor & Francis Group, Boca Raton, London, New York, Leiden (2009)
7. Weintrit, A.: Development of e-Navigation strategy. In: Mikulski, J. (ed.) Advances in Transport Systems Telematics 2. ch. 9 of Section III: Systems in Maritime Transport. Monograph, Faculty of Transport, Silesian University of Technology, Katowice (2007)
8. Weintrit, A.: Common Seas, Common Shores: Development of e-Navigation. Strategy. In: Nincic, D., Benton, G. (eds.) Proceedings of the 9th Annual General Assembly International Association of Maritime Universities (IAMU): Common Seas, Common Shores: The New Maritime Community. California Maritime Academy, San Francisco (2008)
9. Weintrit, A., et al.: Polish Approach to e-Navigation Concept. In: Weintrit, A. (ed.) Advances in Marine Navigation and Safety of Sea Transportation, ch. 10: e-Navigation. TransNav 2007 Monograph, Gdynia Maritime University and the Nautical Institute, Gdynia (2007)

Role of Telematics in Reducing the Negative Environmental Impact of Transport

Jerzy Mikulski[1] and Aleksandra Kwaśny[2]

[1] Silesian University of Technology, Faculty of Transport,
Krasinskiego 8, 40019 Katowice, Poland
jerzy.mikulski@polsl.pl
[2] The General Jerzy Ziętek Silesian School of Management,
Francuska 12, 40952 Katowice, Poland

Abstract. With development of society increase the demand for the movement of people and goods. From years the transportation expand rapidly and is crucial to economic development. Unfortunately, it is also an economic sector with the most harmful effects on the environment. The various modes of transport in different extent impact negative on the environment, which depends mainly on the technical-operating conditions. The up-to-date telematics systems provide the transport industry a revolutionary opportunity to effectively manage. Appropriate use of these systems can help to substantially improve the efficiency of transport and its operating costs, in consequence lead to reduce the negative environmental impact of transport.

Keywords: Transport, transport politics, environment protection, ITS, Eco-telematics.

1 Transport in Environmental Protection Perspective

1.1 Environmental Impact of Transport

Social development is accompanied by the growing need to move and relocate people and goods. Transport, which has been developing rapidly for years, has become a determinant of economic growth. Sadly, it is also the economic sector which has the most detrimental influence on natural environment; its environmental impact has been observed at many levels for years. Main environmental degradation risks are noticeable in the form of:

- air pollution,
- water and soil pollution,
- noise pollution,
- risk of accidents.

The negative environmental impact of transport is manifested on three levels:

- locally – deterioration of the conditions of living, leisure and work,
- internationally – by acid rain or sea water pollution,
- globally – by environmental problems such as global warming.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 11–29, 2010.

Individual branches of transport generate significant adverse effects on the environment, which depend on technical and operational conditions such as: power consumption, structural characteristics, consumable products applied, required infrastructure, demand for a given means of transport.

Globally, the automotive industry emits hundreds of millions tonnes of toxic exhaust fumes per year. Transport ranks very high in terms of greenhouse gas emissions. Allowing for the production of cars, road construction and renovation, transport generates a third of total emissions.

Research studies unanimously point to road transport as the most harmful for the environment. Currently there are 1 billion cars driving on the roads all around the world. It is estimated that the number will have doubled by 2050 [1].

At present, despite considerable effort, greenhouse gas emission levels, in particular those generated by road transport, are still growing. Note that emission level caused by the development of air transport is increasing with equal speed. The increasing availability of jet plane travel, mainly in tourism, results in a huge consumption of fuel, and, consequently, in millions of tonnes of emitted pollutants. Rail transport remains the least harmful for natural environment. Another notable example is water transport, whose effectiveness and haulage capabilities are increasingly growing.

In terms of road transport categories, there is little doubt that cargo transport, especially transport of goods, is the most energy-consuming and environmentally harmful. Its excessive toxicity is exacerbated by bad technical condition of vehicles, numerous cases of overloading as well as damage to road surfaces. Pollution emission is closely connected with considerable consumption of energy by transport.

1.2 Classification of Environmental Risks Due to Transport Activities

Ever year the transport industry releases millions of tonnes of pollutants into the atmosphere. They include lead, carbon monoxide, carbon dioxide, methane, nitrogen oxide, dinitrogen monoxide, carbon compounds and volatile matter, heavy metals (zinc, chromium, copper and cadmium) and solid particles (ash, dust). The extent to which such emission could contribute to global warming is still subject to debate.

The emission of sulphur dioxide and nitrogen oxide is accompanied by the formation of diverse acidic compounds, which – in combination with air vapour – create acid rain, and the precipitation of acid negatively affects the entire environment.

Transport has a significant effect on hydrological conditions. Fuels, chemicals and other hazardous substances released by cars, ships, airplanes and trains in motion or during other activities in ports, airports and terminals may contaminate rivers, lakes and oceans. Since the demand for transport services is on the increase, the emission of pollutants caused by water transport significantly affects the quality of water. Water quality is determined mainly by channel deepening, waste, ballast water and oil leaks.

Waste generated by the activity of ships at sea or in ports is a cause of severe environmental pollution. In addition, the biodegradation of disposed metal or plastic components is very slow, so those parts pollute seas and oceans for many years, making marine navigation difficult.

Ballast is a load element used in ships. Ballast waters accumulated in one region may contain invasive aqueous plant and animal species. Discharging ballast waters in another region may lead to their growth in the new environment and thus disruption to the natural ecosystem of the location.

Oil spills, mainly petroleum leaks caused by accidents or illegal discharge activity, are one of the most severe types of contamination caused by air transport activities. Oil pollution affects marine environment over short distances, but can frequently be propagated by long-range winds and currents, resulting in severe contamination along the entire shoreline.

Degradation of natural environment by transport is caused mostly by the car sector. Soil contamination may occur as a consequence of the emission of exhaust fumes, whereby soils are contaminated by lead compounds and harmful hydrocarbons during petroleum and oil leaks from motor vehicles or the release of cleaning agents directly into the soil. Soil contamination does not affect only areas located in direct vicinity of roads. Research suggests that contamination by harmful dust and exhaust fumes can still be observed over long distances from the road.

Hazardous chemicals, materials and heavy metals are also found in areas adjacent to railways, sea and air ports.

Noise is defined as every unwanted or harmful sound caused by human activity in the open air, including noise emitted by means of transport, traffic, rail traffic, plane traffic and noise originating from industrial activity areas.

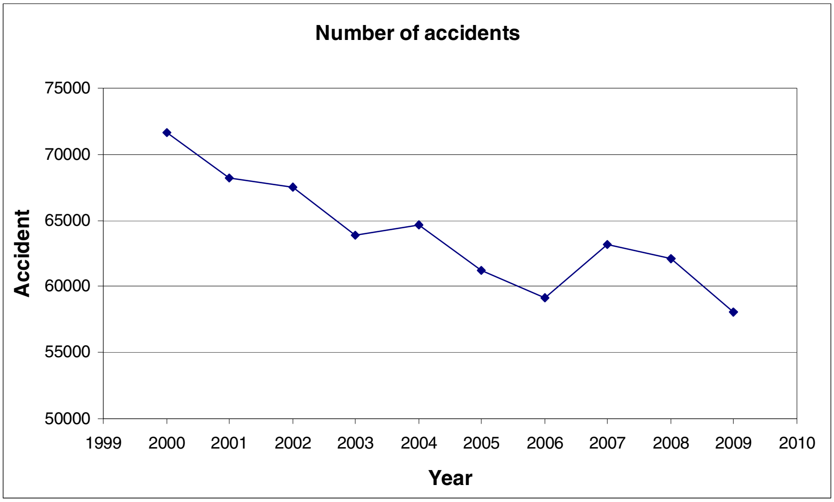
Overall, transport as such largely contributes to noise generation – life in the vicinity of motorways, railways and airports may tremendously influence the incidence of hearing disorders and defects, and due to associated nuisance and harmfulness it has been classified as pollution. It is estimated that one in three EU inhabitants is exposed to excessive noise. According to research, a noise level of approximately 60 dB causes insomnia and fatigue, headache, nervousness, increased blood pressure, mental retardation in children; it also has a negative influence on work and learning efficiency. In case of transport noise it can be said that it creates a scale effect, because intense traffic in main streets and congestion lead to the relocation of transport activities to local streets, which frequently go through residential areas, where increased noise is even more perceptible. Noise also affects animals, distorting their natural ecosystem, frightening them away and often leading to serious accidents.

As a result of the development of transport, the constant need to move and the escalation of overcrowding, the number of transport accidents is increasing. For many years, the highest number of accidents has been reported for road transport.

Each year 1.3 m accidents take place on the EU roads, with 43 dead and 1.7 m injured. Over the last 17 years in Poland there have been almost a million road accidents with a total of more than 110,000 people killed and over a million injured. Each day 15 people die and 160 are injured in road accidents.

According to a police publication on road accidents in Poland, in 2008 there were 49,054 accidents, in which 5,437 people died [2]. However, the number of road accidents decreases year by year. Latest reports inform that in 2009 the number of road accidents was 44,196, with 4,527 deaths, which is over 15% less in comparison to the previous year [3]. Fig. 1 presents the number of accidents in the years 2000–2010.

Road accidents not only cause pain and suffering, but also financial losses paid for by society in the form of national budget expenditure on health and social care. Poland loses 2% of GNP only due to road accidents, since the social cost of a single dead person is almost 1 m PLN.



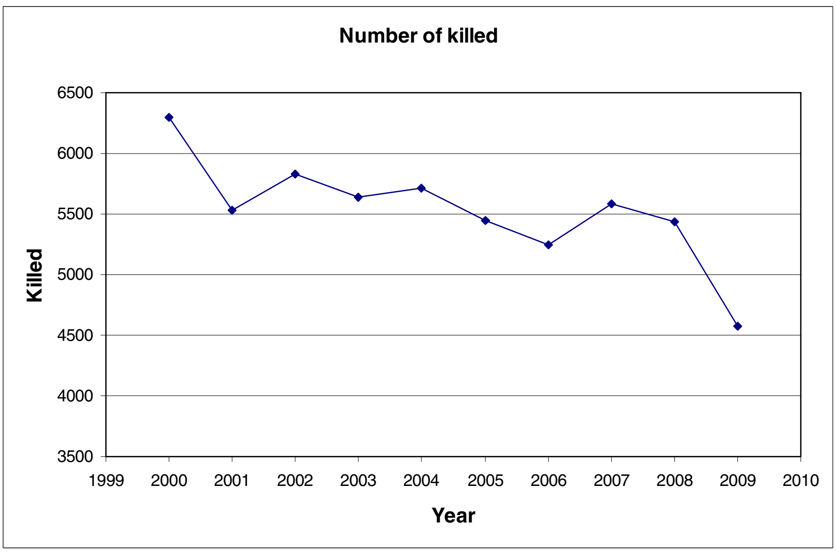


Fig. 1. Number of accidents and number of killed in 2000–2009. Source: own work based on [2].

1.3 Classification of Environmentally Significant Transport Costs

Overall division of costs of transport is the division by cost-bearing entity. In this perspective, the following classification applies:

- internal costs – private expenses paid pay the traveller, company or its owner due to transport activities, and by service providers and their customers,

- external costs – social expenses, generated by the user but paid by the whole society; they are mainly referred to in the context of adverse environmental phenomena caused by transport.

Table 1 shows cost types by cost-bearing entities.

Table 1. Classification of costs and respective bearing entities. Source: own work based on [4].

Cost type	Internal costs (paid by the user)	External costs (paid by society)
Air pollution cost - human health, - animal health.	Generally, users do not pay; they incur their own losses.	Population exposed to air pollution, companies selling health insurance, public authorities.
Climate change costs - natural disasters, - reduced harvest.	Generally, users do not pay; they incur their own losses.	Society, insurance companies, public authorities and future generations.
Noise pollution costs - stress.	Generally, users do not pay; they incur their own losses.	Society exposed to noise.
Accident costs - injuries, - property damage.	Partly as insurance and user's own loss.	Individuals (costs due to pain, administrative costs, etc.), public authorities, insurance companies.
Congestion costs - stress, - reduced efficiency.	Partly by reduced efficiency, costs of user's own time.	Individuals, time loss, companies.
Operating costs	Cost of fuel, infrastructure maintenance, tickets.	Costs incurred by other users and society.
Costs of using infrastructure	Costs covered by charges for using infrastructure, tickets sales.	Costs partly uncovered.

External costs of transport, which are predominately related to the difference between social and private costs, deserve a more profound analysis. They are important due to their connection with environmental activities.

A commonly approved definition says that external costs are generated by undesirable by-products of business activity, do not affect the parties involved and cannot be compensated. This definition also applies to external costs of environmental impact.

In terms of economic analysis, external costs concern the activity of individuals who have a positive or negative effect on other individuals. This concept has considerable significance for environmental issues, since many negative effects are experienced by the entire society In terms of transport activities, external environmental costs need consideration as regards physical damage to natural environment and the assessment of social contribution to such costs. Identifying sources of external costs is a fairly easy task, whereas the assessment of damage and their estimation are yet to reach comparative standards between government and non-governmental organisations.

Table 2. External costs of transport by transport branches. Source: own work based on [5].

	Road	Rail	Air	Water
Infrastructural limitations	Individual transport increases congestion	Timetabled services, making it impossible to provide enough coverage to satisfy demand, different kinds of delays	Timetabled services limit the possibility of travelling outside the schedule and cause time losses	If ports are not filled, congestion occurs individually
Accidents	Loss coverage from insurance, compensation for casualties	Insurance covers most costs incurred by casualties	Insurance covers most costs incurred by casualties	Insignificant
Air pollution	Inhabited areas are located near roads	Costs of generating electricity and internal combustion units	Pollution generated at higher altitudes	Pollution in harbour areas
Noise pollution	Inhabited areas near roads, exposed to noise.	Railway noise causes less nuisance in comparison to other transport branches	More nuisance than in the case of other transport branches	Insignificant
Climate change	Largest greenhouse gas emission	Greenhouse gas emission and costs of generating electricity	Greenhouse gas emissions at high altitudes	Greenhouse gas emission

Therefore, external costs of transport are those who affect society, although they are not directly linked to the user who generated them. They may constitute of the following elements:

- environmental costs (damage due to air pollution, climate change, noise, electromagnetic fields and other environmental consequences),
- uncovered accident costs (grief and suffering),
- congestion (time losses caused by other participants in traffic).

According to other cost types, external costs can be divided into fixed or variable costs. Variable external costs of transport depend on the real use of a given means of transport. Fixed external costs usually appear at earlier or later stages of the transport cycle.

As estimated by the EU, total external environmental costs (excluding congestion) in investigated countries amount to €650 per year, which constitutes 7.3% of GDP of the studied European countries (15 countries of the EU, Switzerland and Norway) [5]. By 2010 it is forecast that without a change in the current European transport policy, total external costs of transport will have increased by 40%.

Studies unanimously demonstrate that car transport generates most external costs of transport.

The distribution of individual cost categories in total external costs is as follows [5]:

- accidents – 24%,
- air pollution – 27%,
- climate change – 30%,
- greenhouse effect – 7%,
- noise – 7%,
- landscape change – 3%,
- the so-called urban effect – 2%.

Fig. 2 presents structure of external costs by type.

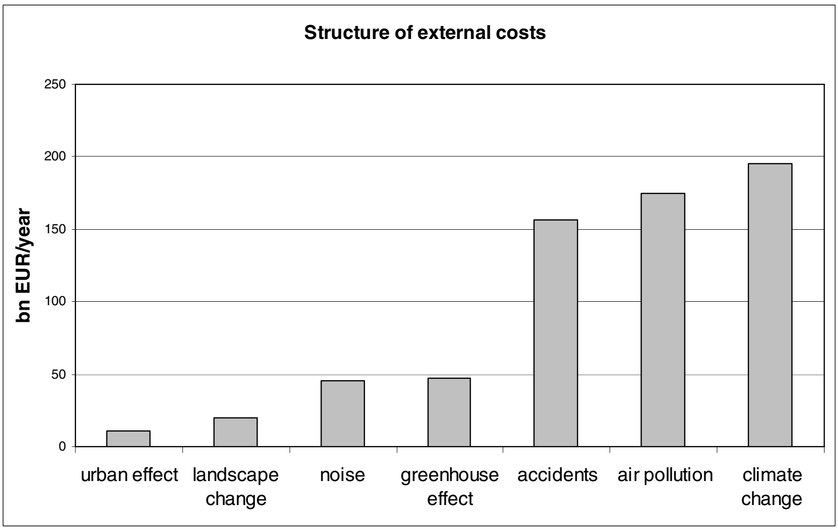


Fig. 2. Structure of external costs of transport by category. Source: own work based on [5].

External cost internalisation is 'delegating' responsibility for the negative impact of transport to users who triggered it. The idea is then to subjectively transport, according to the 'polluter pays' principle. The said subjectification is aimed at fair principles of free market competition, obtaining resources to remedy the negative impact of transport (such as road accidents, air, water, soil pollution, excessive noise, etc.) and focusing on environmental protection as a result of increased awareness of the damage. Main principles of the internalisation of (uncovered) environmental costs are aimed at [6]:

- pollution prevention – transport demand must be satisfied without generating emission which puts public health, biodiversity or the integrity of basic ecological processes at risk and results in global climate change,
- heath and safety protection – transport systems should be designed and operated in a way which protects physical and mental health as well as social welfare, ensures safety of all individuals and enhances quality of life in a community,

- saving natural resources and areas – transport systems must make effectively use of land and other natural resources while preserving natural habitats of living creatures and maintaining biodiversity.

1.4 Poland and EU's Transport Policy in Terms of Minimising the Negative Impact of Transport on Natural Environment

The significance of transport for the global economy is so great that its regulation takes place at all levels, from global through national through municipal.

The fact that transport contributes enormously to the degradation of the environment makes the measures aimed at limiting its negative impact one of the key determinants of economic development. Environmental effects due to transport activities caused EC countries to adopt a common transport policy in order to eradicate the destructive impact of transport on natural environment.

Key Elements of the EU's Transport Policy. Key principles of the European Union's transport policy until 2010 include mainly [7]:

- Sustainable systems:
  - encouraging ecologically friendly branches of transport,
  - integration of means of transport,
  - controlling competition in rendering public services.
- Care for transport system users:
  - improving safety (particularly on roads),
  - appropriate legal service and protection,
  - improving travel quality,
  - transparent charge system.
- Eliminating bottlenecks in architecture:
  - extension or modernisation of transport network,
  - improving traffic conditions,
  - improving the conditions for financing investments in this field.
- Managing transport globalisation:
  - the need to implement global scale programmes such as Galileo,
  - global outlook on transport issues.

Transport as a Component of Sustainable Development Policy. The idea of sustainable development has risen in significance in the recent years. Sustainable development is the fulfilment of the needs of the present generation without impairing the development opportunities of future generations. The notion of sustainable development is based on several elements – the protection of the environment, social development, economic growth and respect for civil rights, all of which should constitute both locally- and globally-integrated process.

Under the common European transport policy for sustainable mobility, a sustainable transport system was defined as a system which should [8]:

- ensure availability of transport goals in a safe way without threatening human health or the environment, equally for the present and future generations,

- enable effective functioning, offer the possibility of choosing a means of transport, support the economy and regional development,
- limit emissions and waste to levels absorbable by Earth, use renewable resources in reconstructive quantities, use non-renewable resources in quantities replaceable by renewable substitutes while minimising the occupation of land and forest areas.

Among quantitative aims of the sustainable transport policy, the following are quoted:

- reducing transport-related carbon dioxide emission, which should not be higher than 50% of emissions in 1990,
- atmospheric aerosol emission level resulting from transport activities should not exceed 10% of emissions in 1990,
- limiting noise to level of no more than 55 dB during daytime and 45 dB at night (outdoors),
- the entire transport infrastructure, including machines for transporting, handling and parking vehicles, should meet air, water and ecosystem protection requirements. Less area should be allocated to transport infrastructure than in 1990.

Among measures intended to support sustainable transport, the following deserve special attention:

- *Introducing clean and effective technologies to transport services*
  The idea is to make transport less traditional and more "intelligent" using advanced technologies, including transport telematics. Measures should focus on improving road safety, including use of advanced active safety solutions in vehicles, limiting emission of pollutants, reducing fuel consumption as well as integrating intelligent transport systems across different transport branches.
- *New road safety technologies*
  Special emphasis is put on the possibility of using state-of-the-art technologies including telematics and ITC, supporting traffic, monitoring and installing devices recording vehicle performance parameters or automatic speed limiters. The European Union also supports launching safe cars fitted with modern safety devices.
- *A real turnaround in combined transport*
  The necessity to look for solutions which could provide an enhanced integration of transport branches and increased interoperability, and oriented at innovative initiatives.
- *Improving traffic conditions*
  The possibility of using telematics when searching for specific traffic management measures, making it possible to precisely manage transport infrastructure.
- *Diversification of energy for transport*
  Actions supporting experiments and promoting "greener" vehicles equipped with hybrid engines, electrical motors or using substitute fuels with lower emission levels.

Principles of Poland's Transport Policy and Their Relation to Natural Environment. In recent years many documents have been devoted to transport policy in Poland. Most important of them include:

- transport policy – an action plan towards transforming transport into a system adapted to the requirements of market economy and new economic cooperation conditions in Europe; a document enacted by the Council of Ministers in 1995,
- alternative transport policy in Poland according to the principles of sustainable development, presented in 1999,
- national transport policy for the years 2001–2015 for sustainable development of the country; a document enacted by the Council of Ministers in 2001,
- infrastructure – a key to development; part of the national economic strategy, enacted by the Council of Ministers in 2002,
- strategies and development programmes for the years 2004–2006 and beyond,
- national transport policy for the years 2006–2025; a document enacted by the Council of Ministers in June 2005,
- strategies and programmes from the years 2006–2008, including a draft of the updated National Land Development Concept.

The objective of the Polish transport policy established in 2005 is to improve the quality and extend the system of transport according to sustainable development principles in the following aspects:

- society,
- economics,
- land development,
- ecology.

In the above mentioned documents the main objective comprises six aims, out of which of special interest are those concerning the improvement of the effectiveness, safety and limiting the negative environmental and health impact of transport.

Among the principles stated in the transport policy, emphasis is put on supporting energy-saving forms and means of transport which create less environmental nuisance and as well as supporting eco-friendly technologies. The focus is also on the necessity of using the opportunities offered by technical development, which provides assistance in organising and managing transport in terms of:

- means of transport, improving effectiveness and safety, reducing negative environmental impact and
- innovative technological and organisational solutions.

As regards means of transport, the role of the state should involve using suitable stimuli to drive the purchase or operation of means of transport having appropriate, desirable performance characteristics and supporting experiments, scientific research and development of new beneficial technological solutions. Advanced technological and organisational solutions, referred to as Intelligent Transport Systems, are also mentioned as the most effective way of supporting transport systems, enhancing their quality and effectiveness, to quote such solutions as:

- advanced road, rail, collective transport, air and vessel traffic management methods,
- fleet management and cargo transport management,
- advanced driver assistance systems,
- automatic toll collection systems,

- automatic control of compliance with regulations,
- dynamic user information systems.

Since the implementation of intelligent transport solutions in Poland is slow, an acceleration is expected in the rate of introducing the above solutions by means of a number of actions, which include:

- designing a nation-wide, long-term modern solutions implementation plan, later also in quantitative assumptions,
- creating national architecture of Intelligent Transport Systems to increase compatibility of transport systems,
- execution of pilot projects pertaining to Intelligent Transport Systems,
- supporting projects related to advanced traffic management systems,
- supporting the development of IT systems enabling the growth of multimodal transport such as cargo tracking features.

Due to the growth of individual transport and the necessity of developing road infrastructure, transport policy assumes the possibility of toughening the climate policy of the EU together with consequences of limiting the rate of the increase of greenhouse gas emission level, supports activities related to the protection of the environment against the negative impact of transport and transport-related activity. In light of the protection of the environment against the negative impact of transport, the following points are of exceptional importance:

- the principle of developing and improving transport system and its branches must be observed by implementing long-term action plans and strategies,
- increasing competitiveness of transport branches other than road and air transport, mainly rail transport, and supporting logistic operators and multimodal transport operators,
- supporting the idea of integrating external cost and eliminating environmentally harmful subsidies with the decision making process, both in relation to transport development directions and transport habits of society and transport companies,
- taking into consideration, to the maximum extent possible, environmental considerations, in particular protecting the environment and building transport infrastructure,
- introducing the obligation to establish transport policy (in the form of statute) on various levels of local government structures, with consideration to the principle of sustainable transport system,
- promoting solutions for spatial and functional integration of transport subsystems,
- publicising best practices and expertise, promoting innovative transport system solutions.

2 Application of Modern Technologies in Limiting the Negative Environmental Impact of Transport

Car manufacturers, research institutes and the automotive industry work on a number of future-proof and modern technologies for vehicles and drive systems. Such technologies unveil the potential of new generation vehicles, which could improve environmental performance levels such as greenhouse gas emissions, exhaust fumes emissions, noise levels and active/passive safety in vehicles. Fig. 3 presents key issues concerning future road vehicles.

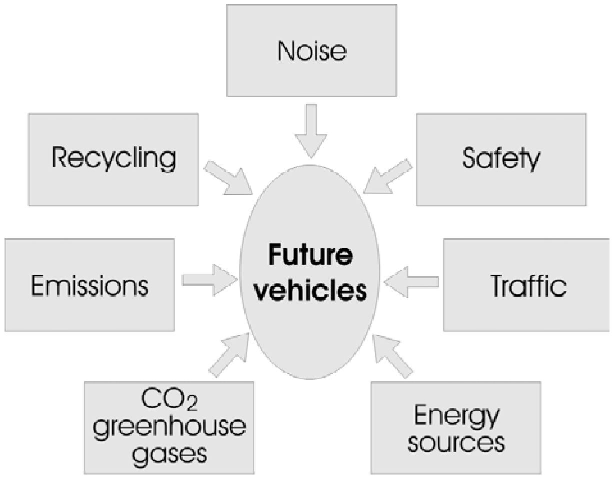


Fig. 3. Key issues concerning technical and performance characteristics of future road vehicles. Source: [11].

In order to influence the future shape of road transport, the global impact of automotive technologies should be taken into account, according to the following categories:

- energy sources and recycling
  - efficiency in manufacturing and recycling processes,
  - economic effectiveness of recycling,
  - future availability of energy sources and their costs;
- infrastructure
  - availability,
  - technological level differences between individual countries and regions,
  - costs;
- environmental impact
  - compromise with future environmental goals,
  - emissions (waste, gas emissions, noise, etc.),
  - presence of entirely new environmental problems in connection with new technologies;
- social and political aspects
  - compliance of new technologies with current and future environmental, social and tax policy,
  - acceptance of new technologies in different cultures and regions (global aspect);

- safety
  - compliance of future technologies with applicable safety standards;
- energy consumption
  - availability,
  - costs,
  - transport and distribution.

All new technologies for classic petrol-powered engines and diesel engines are the first step towards the improvement of future fuel consumption levels and the reduction of greenhouse gas emissions in the near future. When equipped with modern technologies, those engines will have their place on the market until alternative energy sources are made available for mass production. Consequently, telematic solutions for engine control should be constantly developed.

2.1 The Purpose of Telematics and Intelligent Transport Systems

The idea of telematics appeared more than a decade ago and it is possible to define it, in general and simple way, as a communication system for collecting, processing and distributing information.

The transport services market is definitely the most important area for telematic applications. Transport telematics issues constitute a field of knowledge of transport which integrates information technology and telecommunications in applications for managing and controlling traffic in transport systems, stimulating technical and organisational activities which ensure improved effectiveness and safe operation of such systems [9]. Key functionalities of telematic systems include information-operating features. These involve obtaining, processing, distributing information along with its transmission and use in decision-making processes. Such processes encompass both processes managed in a pre-defined way (e.g. automatic traffic control) or processes resulting from dealing with situations as they arise (decisions of administrators, dispatchers, independent infrastructure users such as drivers or pedestrians, etc.) supported by current information [10]. Integrated and cooperating telematic applications constitute Intelligent Transport Systems (ITSs). The basis of Intelligent Transport Systems is to efficiently collect and process information and to manage its flow within the system. This enables supplying information from almost all areas of private and public transport activities in real time.

The characteristics of Intelligent Transport Systems are as follows:

- integration of technology, applied tools and software for efficient flow of information,
- "intelligence" understood as a system’s ability to makes autonomous decisions in varying circumstances,
- flexibility and high adaptability – the possibility of adjusting configuration as necessary,
- effectiveness understood as universality of benefits.

Intelligent transport, supported by a number of integrated telecommunications, IT measurement and control engineering solutions, supported by appropriate tools and software, comprises telematic applications. The have an extensive range of use in many areas of transport, allowing the integration of means and types of transport, infrastructure, business organisation and management processes.

2.2 The Role of Telematics and Benefits Associated with Its Use

The latest telematic systems provide the transport industry with a cutting-edge opportunity for more effective asset management. The proper use of these systems results in a significant improvement of the fleet, a decrease in its mileage, operating costs and fuel consumption, consequently leading to a reduction of the negative environmental impact of transport and increased safety. Note, however, that telematic systems do not form a substitute for a reasonable use of fuel and suitable driver training. Diligent transport operations management, also with the aid of telematics or other technological means, is the key to fleet productivity and effectiveness as well as ecological management. Various telematic applications, depending on their use, produce a number of environmental, functional and financial benefits, both for individuals and for companies.

Transport telematics has an immeasurable impact on natural environment through attempts at achieving maximum use of roads, minimum traffic intensity, assistance in bringing down air and noise pollution and helping to create uncongested traffic zones.

Advantages of using transport telematics:

- time-saving,
- more effective and faster notification of incidents,
- reduction of the number of road accidents and collisions,
- transport cost reduction,
- reduction of environmental pollution and fossil fuel consumption,
- limiting congestion and the number of traffic jams.

Telematic services are constantly developing and will soon become standard features of future cars. Telematics bridges a gap between the vehicle, communications and mobile devices with internet-access. Continuing technological development and the popularity of portable communications devices will make manufacturers try to use telematic services to ensure greater safety, comfort and compliance with environmental requirements.

In order to achieve this, cooperation between car manufacturers and suppliers of telematic services – who will provide open and standard methods for supplying services – is necessary. Next-generation telematics should focus on four basic assumptions:

- **flexibility** of choice between equipment suppliers, call centres, content providers and wireless operators,
- **scalability** – adding services, content and capacity easily,
- **adaptability** in neutral technological platform with normalised interfaces and protocols encouraging the development of new services and easy integration with existing systems,
- **reliability**, which of fundamental importance when providing emergency services for consumers in life-threatening situations.

2.3 Types and Characteristics of Telematic Services

Among key services presently available in the field of transport telematics, the following categories can be listed:

- *Safety and security*
  Safety and security applications include Automatic Crash Notification services, emergency and medical assistance. It is the first set of services offered in the field of telematics. At the same time, these services constituted the basis for the development of the concept of car telematics. Automatic crash notification service monitors a set of sensors in the vehicle. If an accident occurs, information is sent about its severity and location to a centre, simultaneously initiating a telephone call with call centre, where the operator starts to organise suitable rescue operations. If necessary, a request for emergency services to intervene can be sent manually by the driver or passengers. The vehicle unit is frequently equipped with an alternative power supply, so that it can operate also when the car battery or electrical system is faulty. Security features offered by telematic solution suppliers include mainly stolen vehicle tracking, theft alarm notification, remote door control. The unit installed in the vehicle can remotely send periodical information on the exact location of the vehicle, or it can be started automatically by anti-theft sensors fitted in the car. A part of remote door control feature, the unit can also be programmed to open and close the door by remote control.
- *Information and navigation*
  Navigation and information services provide direct access from the vehicle to a number of perfectly integrated features, thanks to which the driver as well as the passengers receive information and content on location. An example of this could include personal access to a network of WiFi/Bluetooth devices inside the vehicle and enabling Internet access by means of a wireless connection. This category also includes a range of services provided for utility vehicles and fleet. Vehicle monitoring and fleet tracking are among services used for improving efficiency and effectiveness of utility vehicles.
- *Diagnostics*
  Diagnostics is another developing area of prospective telematic services. It comprises remote diagnostics, collecting performance data and remote scanning feature. A unit installed in the vehicle is able to perform a detailed diagnostic scan on demand or when certain key threshold values have been exceeded, e.g. the distance covered, previous scanning time, or the time that has passed since the last service.

2.4 The Concept and Purpose of Eco-telematics

For years vehicle manufacturers have focused on improving engine performance and efficiency and have been looking for an alternative drive system in order to reduce exhaust fumes emissions. Meanwhile, for over 10 years ITS technologies have been developing. However, they concentrated mainly on the improvement of safety and comfort of vehicles. Recently their potential for better fuel consumption performance and emission control, also known as "green driving" or "eco-driving" have met with acclaim.

Eco-telematics comprises applications and solutions aimed at reducing negative environmental impact of transport:

- through the driver by providing feedback,
- additionally as navigation and telematic components,
- additionally as drive system components,
- by integrated systems constituting an element of the vehicle's design.

3 Classification of Eco-telematic Applications

There are many telematic systems which may contribute to eco-driving. Generally, they can be grouped according to their role in the driving cycle (Fig. 4).

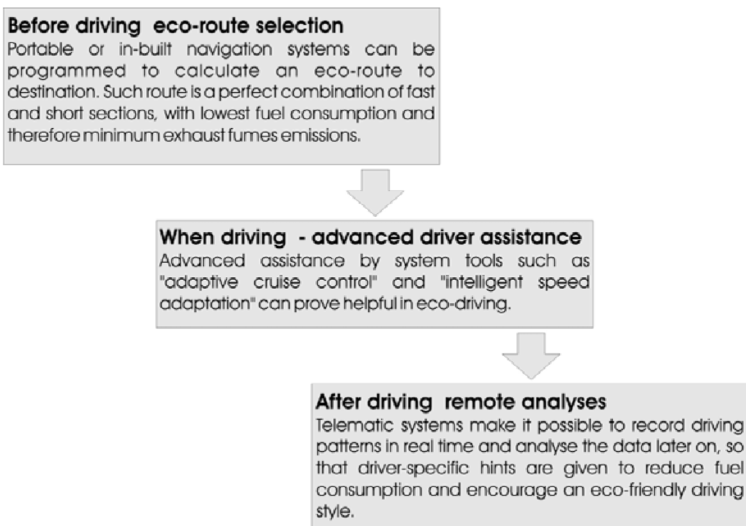


Fig. 4. Division of eco-telematic applications according to the driving cycle. Source: [11].

It is also possible to divide eco-telematic solutions by their function in the vehicle, namely:

- eco-monitoring,
- eco-advice,
- eco-drive.

Fig. 5 presents the suggested division:

- Eco-monitoring
  - eco-compliance,
  - monitoring fuel consumption,
  - monitoring exhaust fumes emissions,
  - monitoring driver's behaviour.

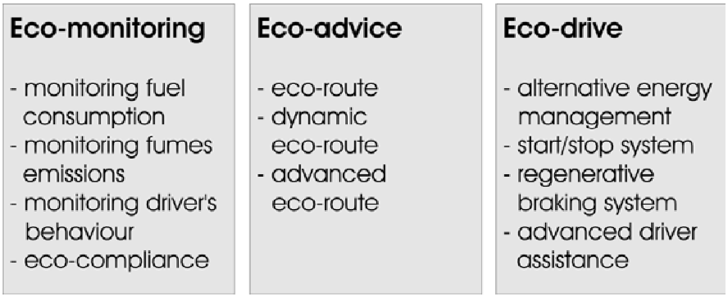


Fig. 5. Suggested division of eco-telematic applications. Source: [11].

- Eco-route
  - route properties integrated in maps: curves, slopes, speed limits, number of roundabouts, etc.,
  - statistical traffic information,
  - possibility of using some vehicle parameters for reducing fuel consumption,
- Eco-drive
  - regenerative braking system,
  - start/stop system,
  - active energy management,
  - controlling the engine by means of the appropriate software.

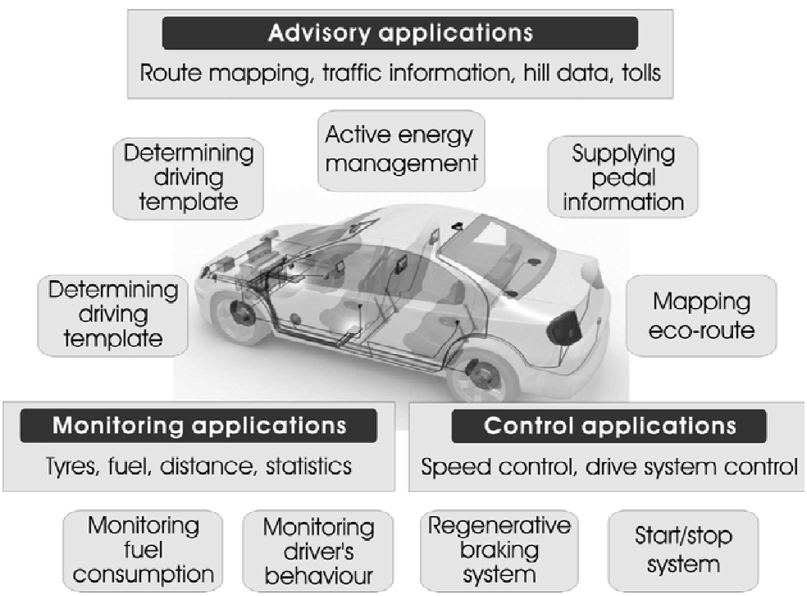


Fig. 6. Eco-telematic applications. Source: [11].

4 Conclusions

Together with growing number vehicles and the development of information technologies, telematics has recently become a "hot" research subject in IT and the car industry. Telematics has gone a long way from providing navigation solutions/assisting the driver to becoming an integral part of the vehicle. Today's telematics ensures safety, comfort and convenience of the driver and the passengers.

Table 3. Eco-telematic applications, their properties and implementation possibilities. Source [11].

	Basic properties	Advanced properties	Implementation possibilities
Eco-monitoring	• On-board driving computer • Monitoring resources: • Fuel consumption • Tyre pressure • Other	• Navigation devices, • Monitoring the ride, • Transmitting data via an Internet website, • Wireless real-time data transmission	• On-board computer is a widespread feature, • Websites are already being set up, • Wireless data transmission 2010+
Eco-compliance	• Remote emission test by means of telematic systems	• Real-time monitoring • Communicating reservations.	• 2011 or later, • state involvement necessary.
Eco-route	Route calculation: • fuel consumption, • emission level.	Route calculation: • traffic information, • weather information, • traffic prediction.	• eco-route mapping is already appearing, • PNDs are already in use today, but will be more widespread in the future.
Eco-drive	Optimising fuel consumption: • engine load, • climate, • air pressure, • other	The use of MAP sensor • An example in Navteq MPE (Map and Positioning Engine Strategy)	• Some systems are already in use.

Explanation of acronyms used in the Table 3:
MAP sensor – a sensor in engine control system is the basic sensor informing of engine load.
PNDs – portable Personal Navigation Devices.

However, due to the environmental protection aspect which is a more and more frequently recurring theme in the field of transport (encouraging eco-friendly, "green" solutions and means of transport), it is quite important to turn attention to the opportunities offered by telematics and Intelligent Transport Systems in this respect. There is a possibility of expanding this vision by other applications, which could directly or indirectly support the protection of the environment against negative impact of the transport industry. Until recently, telematics was associated mainly with navigation services, so focusing its use on transport ecology raises considerable expectations and provides huge space for its development.

Many analyses suggest that telematic applications which allow users to save money will gain popularity in the next few years. Among such applications of special importance are eco-telematic applications which make it possible to economise on fuel.

Both creating and implementing telematic systems is a long-term and enormously complicated process. There are economic as well as social barriers involved, not to mention technological and educational barriers. Creators of telematic services face technological problems, while being fully aware of the need to offer reliable solutions, since in certain situations these solutions will decide on people's life or health. Service purchasers, namely car manufacturers, include additional functionalities in car features, while car owners would like to pay as little as possible. The way forward in this field requires immense effort on part of motor car manufacturers, telematic service providers and governmental administration bodies.

References

1. http://www.hybridcars.com
2. Kopta, T.: Twój samochód zagraża twemu życiu. Polski Klub Ekologiczny, Wrocław (2001)
3. Ministerstwo Spraw Wewnętrznych i Administracji, http://www.mswia.gov.pl/portal/pl/2/8324/Mniej_wypadkow_na_polskich_drogach.html
4. CER Fact Sheet, External costs of transport. The voice of European Railways, Brussels (2008)
5. The New INFRAS/IWW study on the environmental impact of transport. UIC, Brussels (2004)
6. Europejska Agencja Środowiskowa, http://www.eea.europa.eu
7. White Paper European Transport Policy: Time to Decise, Brussels (2010)
8. Commission Expert Group on Transport and Environment, Defining an environmentally sustainable transport system, Luxemburg (2000)
9. Wawrzyński, W.: Telematyka transportu - zakres pojęciowy i obszar. Przegląd Komunikacyjny nr 11, Warszawa (1997)
10. World Wildlife Fund, http://www.wwf.pl/kampanie/kampania_energia.php
11. Kwaśny, A.: Rola telematyki w ograniczaniu negatywnego wpływu transportu samochodowego na środowisko naturalne. Praca dyplomowa, Katowice (2010)

Slovak ETC System Implemented – What Next?

Aleš Janota and Jozef Hrbček

University of Žilina, Faculty of Electrical Engineering, Department of Control and Information Systems, Univerzitná 8215/1, 010 26 Žilina, Slovakia
{ales.janota,jozef.hrbcek}@fel.uniza.sk

Abstract. Electronic toll collection systems represent very efficient instruments for getting fees for using road infrastructure. Despite the widespread of DSRC-based technologies implemented across Europe recently, implementation of satellite-based solutions becomes the crucial issue and pre-condition of future European interoperability in this area. In the context of the new Slovak ETC system put into operation at the beginning of 2010 the challenges of the satellite-based technologies are being discussed. The paper provides an overview of the potential benefits and multiplicative effects achievable through integration with other ITS services.

Keywords: Toll collection, satellite, interoperability, intelligent, transport.

1 Introduction

The idea of paying tolls for using private or public roads and/or other structures such as toll bridges, toll tunnels etc. is known for ages. The first references to paying road tolls are dated to the seventh century BC when tolls had to be paid by travellers using the Susa-Babylon highway in Mesopotamia [1]. For two millenniums after many roads were constructed as toll roads in order to recoup the costs of construction and maintenance. However, rapid development of road transportation in the last century brought a wave of criticism since manually operated toll collection systems required vehicles to stop or slow down which wasted time, raised vehicle operating costs and had negative effects on environment. Renewed interest in toll roads in last two decades has grown mostly for two reasons:

a) Toll collection provides a mechanism for financing construction of new and maintenance of existing roads.
b) Technological innovations are increasingly making possible an effective use of time-of day pricing on toll roads to control and mitigate problems of congestion.

Technology change has been transforming tolling in the last two decades and offers huge possibilities for improving further into the next 20 years horizon. The need for a change together with new technologies and financial sources has brought an innovative solution of the new concept called electronic toll collection (ETC) where no toll booths are required any more. Generally, innovation is the process of turning ideas into manufacturable and marketable form. The electronic system determines whether a passing vehicle is enrolled in the program and alerts enforcers if it is not. The tolls

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 30–37, 2010.

are collected by a scheme that does not make the vehicles to stop, nor to reduce speed and nor to change lanes. Despite of continuing criticism claiming that “tolls are a polite form of a highway robbery” this concept has been growing rapidly in recent years and represents one of the most successful applications of the intelligent transport systems (ITSs).

From technological point of view, there are several key technology features that drive the transformation. The first deployments of the ETC were based on Dedicated Short Range Communication (DSRC) transponders or tags or On-Board Units (OBUs). Some systems have been based on image capture through video technology. The toll mechanism in vehicle positioning systems is based on interaction between OBU and backend system through mobile network instead of communication with Road Side Unit (RSU) as in the DSRC-based ETC systems. The systems adopt the General Packet Radio Service (GPRS) mobile network and the Global Navigation Satellite System (GNSS) as the basis. The GNSS technology within OBU estimates position by combining measurements of signals from a constellation of orbiting satellites, typically Global Positioning System (GPS) or the Global Orbiting Navigation Satellite System (GLONASS). Cellular Network technology is used for bidirectional communication between an OBU and a fixed network of terrestrial transmitters or mobile transmitters (Global System for Mobile Communications - GSM) [2], [3]. These technologies may also be applied integrated with odometer and other inertial sensors.

In January 2005 a satellite-based system launched successfully in Germany and in January 2010 in Slovakia. However, “too rapid” development of these systems has brought the actual problems with interoperability since the real implementations have got ahead of standardization process, mostly represented by the CEN/TC 278. Considering initial several month experience with the Slovak satellite-based system the paper has been written to indicate and foresee future benefits and multiplicative effects mostly achievable through integration with other ITS services. ITSs are used to improve traffic flow, to increase the efficiency of freight and public transportation and to reduce fuel consumption. They also have been identified as a tool to improve road safety.

2 ETC Potential for Future

The ETC system considered as an isolated autonomous system will bring only limited number of effects and benefits. Therefore one should look into a future in connection with cooperation with other ITS-based technologies, in the context of so needed interoperability and under the present conditions that are to be changed: road congestion leads to economic losses of about 1 % of the GDP, each year still more than 40,000 citizens are killed on the road and road transport accounts for 72 % of transport-related CO_2 emissions (transport is the only sector where CO_2 emissions still grow, for period 2010-2020 the growth +15 % is estimated). Implementation of the interoperability of ETC systems is included in the ITS Action Plan as one of key priorities within the 2nd priority area “Continuity of Traffic and Freight Management” [4]. According to [5], electronic fee collection is expected to be used in Europe by nearly half of all vehicles (about 46 %) by 2020, compared with 3.7 % in 2005.

2.1 First Observations from the Slovak ETC Implementation

From January 1, 2010, the Slovak Republic changed the system of charging the road infrastructure from highway stickers to the ETC system. The change has applied to motor vehicles with the total weight over 3.5 t, vehicle trains with the total weight over 3.5 t determined for the transportation of goods, motor vehicles, enabling the transportation of more than nine persons including a driver, except for vehicle trains created by a motor vehicle of the categories M1, N1, M1G and N1G [6]. The toll is paid for the use of specified sections of highways, expressways and the 1st class roads. Legal Regulations of the Slovak Republic in the field of ETC include three Acts[1], two decrees[2] and one government regulation[3]. They have taken over two Directives of the EC and EU, too [7], [8].

After launching the system many objections were raised by hauler representatives criticising its unpreparedness (absence of foreign language manuals, bad price calculations, ambiguous definition of tolled sections, exactness of position localisation, etc.). However, during the first 13 days almost EUR 1.1 mil. was collected. Rising campaign against the system resulted in breakdown of the system for 46 days from 14th January till the end of February (pecuniary injury was estimated between EUR 6.9 mil. and EUR 9.2 mil.). In the first four months of operation the complex service brought the amount of almost EUR 38 mil. for the state. As at 30 April 2010 as much as 166,328 vehicles were registered in the toll system. As much as 80 % of them belonged in the category over 12 t. On average the specified road sections were used by 27,416 vehicles per day. Of the total amount of the collected tolls, over 43 % were paid by foreign haulers. In most cases (90 %) they chose the prepaid regime unlike ca 65 % of vehicles operated by the Slovak haulers who used that payment regime [6]. Expected income from the ETC system in 2010 is ca EUR 180 mil.

2.2 A Need for Assessment of New Solutions

Whatever new transportation solutions will be implemented in near future with a view to get new functionalities of the ETC system, their impacts should be assessed carefully. Little attention is usually paid to the interactions that may occur among multiple measures despite they might have consequences that are more significant than the

---

[1] Act No. 25/2007 Coll. on Electronic Toll Collection for the Use of Specified Sections of Ground Roads and on Amendments and Supplements to Some Laws in its Valid Wording (2007).

Act No. 725/2004 Coll. on the Conditions of Vehicles Operation in Traffic on Ground Roads and on Amendments and Supplements to Some Laws as Amended (2004).

Act No. 639/2004 Coll. on Národná diaľničná spoločnosť in its valid wording.

[2] Decree No. 108/2010 Coll. (Toll Order) of the Ministry of Transport, Post and Telecommunications of the Slovak Republic governing details of Toll Collection According to the Empowering Provision of the Act No. 25/2007 Coll. in its Valid Wording (2010).

Decree No. 61/2010 Coll. of the Ministry of Transport, Post and Telecommunications of the Slovak Republic Specifying the Sections of Highways, Expressways and the 1st Class Roads with Electronic Toll Collection (2010).

[3] Regulation of the Government of the Slovak Republic No. 350/2007 Coll. Stipulating the Toll Rate Amount for the Use of Specified Sections of Ground Roads (2007).

impact of any one individual measure. One of possible assessment approaches is based on considerations of supply-demand relationship. It may be readily visualized using traditional supply-demand curves as shown in Fig. 1. Usually there are four classes of measures identified, heaving different impact on the analysed relationship [9]. There are measures of the 1st class that represent actions reducing travel demand, for example tolls, fares etc. Measures of the 2nd class represent actions that enhance highway supply, for example construction of additional lanes, better signal timing etc. Measures of the 3rd class are actions that reduce both demand and supply, for example taking a lane for high occupancy vehicles. Measures of the last 4th group represent actions that reduce demand and enhance supply, for example new high occupancy lanes. Each group of measures causes shift of one or both curves in a certain direction. As a consequence the equilibrium point is moved, representing improvements in mobility, fuel consumptions and emissions. Classes of measures 1, 2 and 4 improve mobility (increased supply by the downward movement of the supply curve, and increased demand by a shift to the right of the demand curve). The class 3 reduces demand, but offers the potential benefits of reduced fuel consumption and reduced emissions. More detailed description of this assessment approach is available in [9].

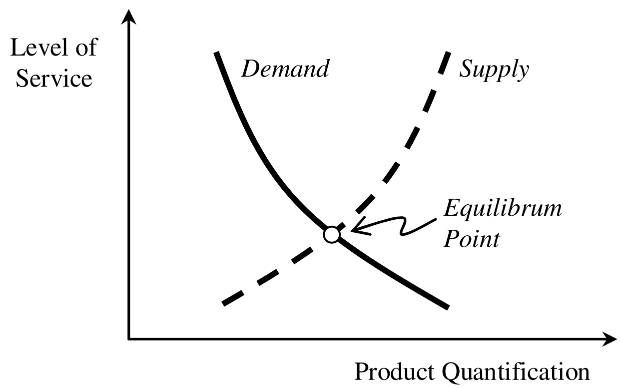


Fig. 1. Supply-demand relationship: the vertical axis represents level of service (expressed in terms of increased travel cost, trip time, user cost, etc.), the horizontal axis represents a product quantity (e.g. expressed in terms of vehicle kilometres of travels etc.)

2.3 Could Electronic Toll Collection Be Even More Fair?

Apparently, there is no need to dispute whether toll collection is needful or not. However, we could discuss whether it is sufficiently fair. Probably the only correct criteria applicable when answering the indicated question is the fact how the collected financial means are finally spent and what added value they bring to the society. Reasons given for ETC implementation are usually as follows:

a. Getting financial means for maintenance, reconstruction and construction of new sections of the road network;
b. Getting financial means for delivering services related to operation on road communications;

c. Possibility to pursue environmental targets (preferences of those traffic means that produce less pollution);
d. Harmonization of conditions for usage of communications by different transport means, or different haulers;
e. Rationalization of usage of the highway network, network of expressways and roads of the 1st class, especially by freight service;
f. Acquisition of a new instrument for strategic and dynamic control of traffic flows state-wide;
g. Acquisition of a new instrument for increasing transport safety.

At present, the Slovak ETC system uses the most common practice – setting a fixed toll rate based upon vehicle characteristics such as the number of axles. Thus it applies distance charging when the travelled distance in a toll road network is considered together with a certain category of the vehicle representing how much a vehicle damages the road infrastructure and generates environment pollutions. On the other side, the real weight of the vehicle depends on the actual load. Though the actual state of technologies would make real weight measurements possible, not much attention is paid to this fact so far. Absolutely, no examples of measuring real weights for the purpose of ETC calculations are known in the world yet.

In addition to tolling based on distance charging, there is also potential for use of time and location charging. Importance of both of them will rise together with a need to solve congestion problems [10]. The first practical usage of time charging in the region of Central Europe can be observed in the Czech Republic where DSCR-based ETC system has been in operation since January 1, 2007. The public generally easily understands time-of-day schedules. Peak period charges or variable toll rates based on time of day are already in common use in industry. Another advanced level could involve dynamic or traffic conditions-based pricing. By using traffic sensor information, real-time traffic conditions could be determined and used to update prices as conditions change [11].

2.4 ETC as an Economic Instrument for Managing Congestions

Traffic congestion in Europe is bad and getting worse. Because of a lack of proper investment in road capacity, congestion costs Europeans more and more every year. On one side we use more and more fuel-efficient vehicles which allow us to reduce gasoline consumption per km driven. On the other side the revenues from gas taxes drop and the need for road construction and maintenance continues to grow. In this context adoption of the doctrine „pay-as-you-go“ and its embedding in the tax-and-grant system seems to be the best step. It is estimated that ITS technologies could potentially reduce congestions by 5-15 % at least. Namely ETC systems, dynamic navigation and dynamic traffic and freight management systems seem to be most prospective in solving that task.

Population growth goes along with an increase in car ownership and demand for transport activities across all regions. The world is urbanizing rapidly, and population densities are increasing. A United Nations report estimates approximately 70 percent of the world's population will live in cities by 2050 [12]. In several countries there are designed special lanes reserved to vehicles carrying at least two users (high occupancy lanes) to help move more people though congested areas. Detecting the number of passengers in vehicles can be useful to grant access to or to control vehicles in these lanes. Variable highway toll fees depending on passenger number and traffic conditions (high occupancy toll) is another possible application.

2.5 Environmental Effects

Transport is a major consumer of fossil fuels, currently accounting for 30 per cent of total energy consumption and 28 per cent of CO_2 emissions in Europe. Traffic accidents and congestion are costly to society in terms of lost lives, productivity, and energy. For transport to become more efficient, safe, and environmentally sound, new ways of looking at overall transport objectives are needed. To reduce congestion and make public transport more attractive means to reduce transport-generated pollution, encouraging conditions for sustainable economic growth. Generally, greening of road transport covers three main domains - Electronic Toll Collection, Navigation and eco-driving and Green transport corridors. Currently, 3-4% of the world-wide energy is consumed by the ICT infrastructure (approximately, 1.5%-2% by wireless), which causes about 2 % of the world-wide CO_2 emissions (which is comparable to the world-wide CO_2 emissions by airplanes or one quarter of the world-wide CO_2 emissions by cars). One of emerging topics being discussed in recent years concerns a future trade with personal emission quotas. The idea is based on monitoring of individual car operation, its measuring, scoring, and keeping files of „individual emission account".

Using taxes and fees to control wasteful driving habits while helping the environment often results in the controversial debate. However, practical examples such as the Stockholm Congestion Charging System launched a few years ago have shown how the congestion charging system can reduce traffic volumes, decrease CO_2 emissions and improve accessibility and bring significant benefits to the city, its visitors, and residents [13]. The Stockholm system is the largest of its kind in Europe, with 18 barrier-free control points around the inner city equipped with cameras to identify vehicles around a 24 square kilometre area.

2.6 Additional Services of the ETC System

The ETC system can offer a lot of additional services in the field of after-theft measures, emergency and safety systems, supervisory and monitoring systems, traffic management and payment for services. GNSS based after-theft (or anti-theft) systems are equipped not only in high value but also normal vehicles. The system can track location of vehicles and raise the awareness of theft. Special vehicles may also be disabled by remote control. Safety and emergency systems involve e-Call service in case of accident, automatic warning systems monitoring road surface conditions, in-vehicle systems warning against pedestrians ahead. Supervisory and monitoring systems are oriented to measurement of section and/or spot speed, monitoring of traffic and weather conditions, checking tax and other fee payments, monitoring of hazardous and abnormal goods movement, vehicle height check before entering a tunnel, fleet managements for the needs of haulers, smart traffic signs communication with OBUs, access control for authorized vehicles entering a protected area, and vehicle weight on the run. Traffic management helps to get traffic analyses and characteristics, perform automatic excess identification and automatic detection of traffic conditions (based on according to travel times). ETC systems can also be integrate with collection of other fees (parking, fuelling, etc.).

A large amount of data cumulated by the ETC system could be further used in other services. “Pay as you drive” approach has the potential to change policy of insurance companies since sensitive data on vehicle operation might be available (what distances are usually moved in a highway network, network of expressways, the 1st class roads or other roads; at what parts of day and on which days is the vehicle mostly used; what is the share of urban and country road usage etc.). If the vehicle operator does not use the vehicle for a certain period (e.g. because of vacation) the insurance tax should/could be lowered.

3 Conclusions

The use of ETC system beyond its primary purpose of automated toll collections enables toll system operators to save costs, increase road safety and public security, improve service for road users and generate additional revenues through value added services. The authors summarized the first observations from operation of the Slovak GNSS/GPS electronic toll collection system and showed what further services could be implemented in near future.

Acknowledgements. The paper was prepared under the partial support of the ERDF - European Regional Development Fund, project No. ITMS-26220120028, “*Centre of excellence for systems and services of intelligent transport”* and partial support of the Slovak Grant agency VEGA, grant No. 1/0023/08 *Theoretical apparatus for risk analysis and risk evaluation of transport telematic systems*.

References

1. Gilliet, H.: Toll roads – the French experience. Transrouts International, Saint-Quentin-en-Yvelines (1990)
2. GSM Coverage Maps, `http://www.gsmworld.com/roaming/gsminfo/index.shtml`
3. Lee, W.H., Tseng, S.-S., Wang, C.-H.: Design and Implementation of Electronic Toll Collection System Based on Vehicle Positioning System Techniques. Computer Communications 31, 2925–2933 (2008)
4. The ITS Action Plan, COM, 886 (2008), `http://ec.europa.eu/transport/its/road/action_plan_en.htm`
5. Intelligent Transport Systems. A Smart Move for Europe. European Communities (2009), `http://ec.europa.eu/transport/publications/doc/2009_its_factsheet_en.pdf`
6. Mýto, `http://www.emyto.sk`
7. Directive of the European Parliament and Council 2004/52/EC of April 29, 2004, on Interoperability of Electronic Road Toll System in the Community (2004)

8. Directive of the European Parliament and Council 1999/62/EC of June 17, 1999, on the Charging of Heavy Goods Vehicles for the Use of Certain Infrastructures (1999)
9. Tarnoff, P.: The Thinker. Thinking Highways. Europe/Rest of the World Edition 1(1), 26–29 (2006)
10. Samuel, P.: The Role of Tolls in Financing 21st Centrury Highways. Policy Study 349, Reason Foundation (2007)
11. Toll Technology Considerations, Opportunities, and Risks. Washington State Comprehensive Tolling Study, Final Report – Volume 2, Background Paper #8 (2006)
12. Handwerk, B.: Half of Humanity Will Live in Cities by Year's End. National Geographic News, http://news.nationalgeographic.com/news/2008/03/080313-cities.html
13. Houghton, J., Reiners, J., Lim, C.: Intelligent Transport: How Cities Can Improve Mobility. IBM Research Report GBE03232-USEN-00 (2009)
14. Mikulski, J.: Legislation changes on introduction of toll collection system in Poland. In: Conference proceedings of International Symposium on Stochastic models in reliability engineering, life sciences and operations management. Beer Sheva, Israel (2010)

Technological Support for Logistics Transportation Systems

Andrzej Bujak, Zdzisław Śliwa, and Alicja Gębczyńska

Wroclaw School of Banking, Faculty of Finance and Management,
ul. Fabryczna 29-31, 53-609 Wrocław, Poland
sliwazd@wp.pl

Abstract. The modern world is changing introducing robots, remotely controlled vehicles and other crewless means of transportation to reduce people's mistakes, as the main cause of incidents and crashes during traffic. New technologies are supporting operators and drivers, and according to some studies they can even replace them. Such programs as: AHS, UAH, IVBSS or MTVR are under development to improve traffic flow and its safety, to reduce traffic hazards and crashes. It is necessary to analyze such concepts and implement them boldly, including Polish logistics' companies, new programs, highways' system etc., as they will be applied in the future, so it is necessary to prepare logistics infrastructure ahead of time in order to capitalize on these improvements. The problem is quite urgent as transportation in the country must not be outdated to meet clients' expectations and to keep pace with competing foreign companies.

Keywords: logistics, telematics, transportation systems, logistics services Automated Highway System, logistics' technologies.

1 Introduction

The contemporary world is changing rapidly introducing robots, remote controlled vehicles and other crewless means of transportation aimed to reduce all the people's mistakes as the predominant cause of incidents and crashes. Moreover, traffic congestion is causing demands for new solutions to improve transportation reliability and the whole transportation system both for passengers and goods movement along highways and cities. For logistics, the transportation is a kind of bloodstream, which is connecting all the participants of the system, producers and final users, even continents and as such is the vital and fragile element. As a result, undisturbed traffic flow and safety are key enablers of an effective transportation infrastructure and system suitable for future needs. Safety issues are mainly the result of human errors, as "ninety percent of today's accidents are caused at least in part by drivers; about 70 percent are caused predominantly by human mistakes and 20 percent have some kind of component of human error that influenced a cause of an accident" [6]. To avoid them, new technologies are increasingly supporting transportation operators and drivers; according to some new research, they may even replace them. Such solutions are under quick technological development mainly in developed countries, as they are

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 38–50, 2010.

connected with high cost, technological capabilities and research and development infrastructure. The subject is not new in Poland and many institutions are following the development of such concepts including institutions of higher education, transportation companies and governmental agencies. Among them, the Wroclaw School of Banking is closely following such new trends, as according to new experiments, they will be introduced into logistics transportation systems [8]. Such concepts will be implemented in the transportation structure within 20-30 years as the problem is rather complicated, but the outcome could be profitable for logistics' providers, especially as long-term investment. Such concepts are under constant development in many countries with the USA in a leading role. Moreover, they have both civilian and military applications. Even now, some automated robots and vehicles have been effectively implemented encouraging civilian society and scientists to continue progress in this area. However, not all institutions are able to conduct such experiments as infrastructure and vehicles' systems necessary for tests are expensive, so big investments are needed for progress. As a result, smart highways and new type vehicles equipped in sophisticated sensors will be put into operation basically in developed countries with strong and rich transportation companies, ready to fund research and infrastructure. There are many future related concepts aimed to improve transportation systems. Among them the Automated Highway System (AHS) and Underground Automated Highway (UAH) have been developed and they are under serious research, as they could have promising effects on traffic management. Moreover, they are directly related to employment of enhanced trucks equipped in modern sensors to support the system and people and even crewless vehicles.

The aim of the paper is to present current projects connected with improvement of transportation as an essential part of modern logistics and to elaborate on the main directions of such research. First, the support for the drivers will be discussed, which will be followed by covering new concepts of constructing future highway systems. Next, the programs related to autonomous vehicles will be under study. Such research will be based on case studies, which can present new and broad system wide approaches demanding a comprehensive approach of scientists, national and regional level official organizations, working hand-in-hand with logistics transport companies. What is unifying these partners is the common understanding of the need to improve transportation as vital part of logistics along with their will to invest funds in the short term to realize profits later.

1.1 New Solutions to Support the Personnel

New concepts are putting an accent on supporting people to enable better performance during their daily duties. So, such advanced ideas will directly support system operators and drivers. At the same time these ideas will indirectly positively influence logistics companies by improving their effectiveness, safety and reducing costs. They are under constant research especially in developed countries as they are connected with huge costs, modern technologies, high level academic institution involvement and appropriate technical culture. One of the very important issues connected with technological support for logistics transportation is security, which is directly influencing the continuous flow of logistics supplies. There are many projects to improve the situation, which are conducted in almost every developed country, based on regional circumstances, capabilities and needs. Some of them are international in nature connecting educational institutions, national level organizations and also logistics companies to combine funds providers, scientists and final users of new technological solutions. Among many current projects, a very interesting example there is the cooperation of the Transportation Research Institute of the University of Michigan (UMTRI), U.S. Department of Transportation (US DoT) and their automotive and commercial vehicle industries' partners: Visteon Corp., Eaton Corp., Honda R&D Americas Inc., International Truck and Engine, TK Holdings, Battelle, Con-way Freight. Appropriate composition of participants is a guarantee of a very serious approach to the research as every organization is really concerned about the final outcome of the inquiries. The number of the participants also provides a comprehensive approach to solving the problem, which is critical as merging practical experiences and theory is supporting the process to improve transportation with reliable and long-term solutions. The total value of the agreement is about $ 32 million and is focused on "developing and testing a new, integrated crash warning system in a fleet of 16 passenger cars and 10 heavy-duty trucks" [5] in the frame of the Integrated Vehicle-Based Safety Systems (IVBSS) concept.

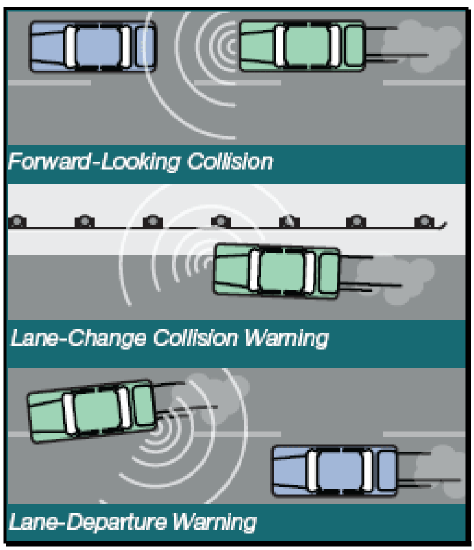


Fig. 1. The main road situation researched in IVBSS concept. Source: [17].

Regarding the conceptual phase of the project and research, the university is the essential contributor to the agreement, which is also important as it has all necessary capabilities to develop proper methodology and theoretical background of the project, and later is able to coordinate and to carry out all required laboratory and field tests. The concept is also a comfortable and reliable solution for the other contributors, as their involvement is rationalized based on real needs and complex approach needed to reach their expectations. The main focus of the program is to develop "integrated, advanced technologies that can help drivers avoid crashes" [5]. The research covers dangerous situations for drivers on highways especially when they are: leaving a highway, changing or departing lanes, a risk of collision with a car in front of them or if they are approaching a curve at excessive speed (see Fig.1 and Fig.2). These solutions, supported by inertial, video, radar sensors and GPS modules, will deal with some 67% of main situations leading to collisions, and those situations are rather an important factor causing difficulties for land logistics transport.

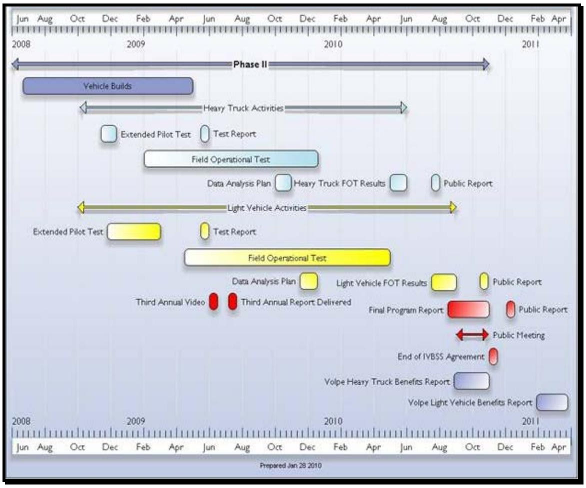


Fig. 2. IVBSS Program Roadmap. Source: [5].

In 2008 and 2009 the project was focused on creating and verifying prototype systems, which later were heavily tested to ensure their safety for field trials and meeting basic requirements of the concept. In 2009 field tests were performed. In the case of the heavy trucks (IVBSS-equipped International ProStar 8600-series trucks), operating as a part Conway Freight, Inc. car fleet, 20 volunteer drivers were using them for regular duties over the ten month period. The total driving time reached 16500 hours, covering some 650000 miles (over 1 million kilometers), enabling collection of full travel data for 140000 miles (224000 km), and the rest was performed with all the systems operational. Such tests had one more important feature as they had been performed in the real operating environment including: infrastructure, employed drivers and all-weather conditions. The results and conclusions were presented to the US DoT and business partners by UMTRI in November 2009 as "*Integrated Vehicle-Based Safety Systems (IVBSS) Heavy Truck Platform Field Operational Test. Data Analysis Plan*" [9]. The results are rather promising and at the end of the research the findings will be implemented in designing future integrated systems for vehicles to improve safety, efficiency and reliability of transport for the benefit of logistics providers. Such cooperation is a very good example of effective, focused, comprehensive and joint teamwork of a few varied organizations to support technologically transportation systems. In addition to profit potential this research is also providing other advantages to the companies involved over other transportation providers. For academic institutions it is a chance for their development by getting funds to do it. Not only is the

USA concerned regarding new solutions, also the European Union is heavily supporting transport and logistics improvements "*through research and development projects, thematic networks, concerted actions and integrated projects*"[16] as the importance is clearly understood to be competitive and effective in the future.

1.2 Current Driverless Concepts Development

As mentioned before, quality people are the most important resource for any system, but simultaneously they are also the main reason of incidents and malfunctions. As a result, some of new concepts, which are under research, are connected with eliminating people's mistakes by using driverless vehicles, which is also connected with preparing appropriate infrastructure to support them. They are promoted in the frame of the Intelligent Transportation Systems (ITS) and are under research conducted by many well-known academic institutions. For example the Defense Advanced Research Projects Agency (DARPA) organized Urban Challenge in November 2007 at the former George AFB in California as a continuation of the 2004 and 2005 Grand Challenges. The main purpose was to build an autonomous vehicle capable of driving in traffic, performing complex maneuvers such as merging, passing, parking and negotiating intersections. At the beginning 89 teams applied for the competition. Such initiatives are important step to speed up the implementation of new driverless vehicles into traffic and transportation systems, logistics and even for military logistics purposes. During the competition, constructors met three main technical challenges [1]:

- to install drive-by-wire technology in the vehicle and to modify the vehicle to fulfill DARPA safety requirements including equipment to provide electrical power for computers, actuators, and sensors,
- to provide sensing and information fusion algorithms as the autonomous vehicle should be able to sense both vehicle current status, such as: position, speed, direction, environment information, the existence and position of obstacles,
- to provide full control of vehicle based on the fused information, the autonomous-driving system to make the correct decision, apply the navigation algorithm, and properly control the vehicle through the driveby- wire capability. Visual road detection, map-based path finding, and real-time communication between computers were also important issues.

Such competition is evidence that ITS concepts are very important and real future solutions for different purposes and they are naturally enforcing the competition among participants. At the same time, the creativity is not hampered, allowing a search for new nonstandard ideas. Logistic requirements will be very important aspects of new capabilities provided by such innovative concepts. Ohio State University, in collaboration with Oshkosh Truck Corporation, had created Team TerraMax, which constructed an MTVR (Medium Tactical Vehicle Replacement) truck for Grand Challenge 04 being completely autonomous vehicle in order to support military requirements.

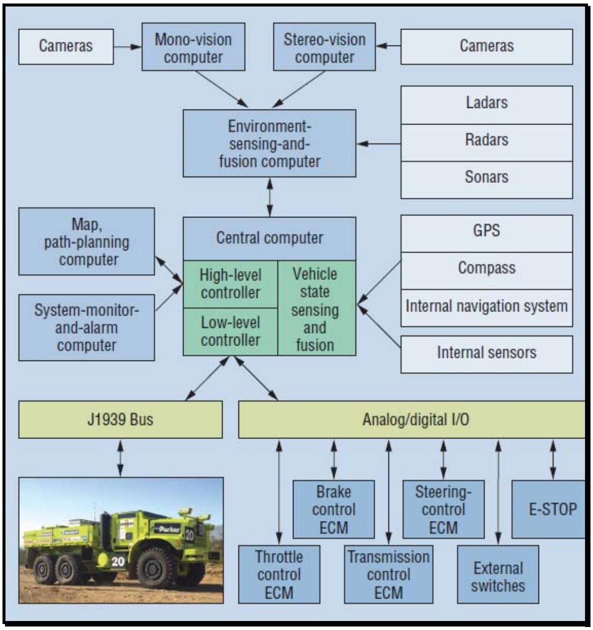


Fig. 3. TerraMax's hardware (example). Source: [1].

TerraMax hardware included "*the drive-by-wire electronic actuators, local network, computers, vehicle ego state sensors, and several different types of environment sensors*"[1] (see figure 3), including six computer systems for navigation and control. All the connections "*were virtual links and the connections between computers were linked through a local network, and the connections between computers and sensors via hardware interfaces*" and "*directly connected analog and digital signals*"[1].

Experiments revealed the necessity to upgrade and stabilize the network and inter-process communication especially in more complicated and demanding environments, like built-up areas and crowded roads [1]. So, new sensing technologies including scanning radars are under research along with tests to place sophisticated systems on smaller platforms. These means have the potential to create driverless capabilities for a wider range of logistics services by reducing hazards and crashes, although they are connected with new risks, uncertainty and the uncertain performance in changing circumstances.

1.3 New Infrastructure Solutions

To fully operate such systems it is necessary to improve transportation infrastructure especially highways, which should contain many characteristics to simplify new challenges caused by automation, including uninterrupted traffic flow, new type crossroads, controlled access for vehicle, and so on. The present status of highways will be significantly useful to implement new transportation solutions, for instance the problem of crossroad collisions can be reduced by "*activating the onboard warning systems and automatic braking systems with electronic signal lights in addition to the* *normal traffic signal. If the intersection detects a potential for a collision it can notify equipped vehicles*" [12]. Moreover, one of challenges during smart vehicle deployment on automated highways will be "*integration of cyclists and motorcyclists to the Automated Vehicle Control (AVC)*" [10] systems. In Parallel, the effects of both pedestrian and animal traffic will need thorough research.

Another challenge will be connected with overall acceptance of new systems which are driverless, privacy issues related to Automatic Vehicle Identification (AVI) along with Automatic Vehicle Location (AVL) concepts, and others. The issues are heavily connected with the unpredictability of traffic situations and the best way to cope with them; as a result a successful AVC system must address many vital issues, like false alarms and system failures. If they will be overcome such challenges acceptance will come rather soon, but accidents and failures can ruin public confidence in very short time. So, the challenges are quite substantial and the systems must be checked very thoroughly before entering daily traffic. Additionally, there are some countries with very poor infrastructure and very bad driving habits, so new concept implementation will be, in some cases, restricted to only smart roads, and the cost-benefit ratio will be rather unacceptable. The concepts mentioned before have worldwide significance as every country has similar problems and only close cooperation will enable joint solutions. As a result, many international conferences have been held share experiences. According to Chuck Thorpe, who runs the Navlab group in the Robotics Institute of Carnegie Mellon University, "*we need cars that are much more intelligent than the cars that we have today. This is not because we want to take people out of the cars, but because we want to help people get to where they need to get too much more safely and much more efficiently than we can today. People have done a good job driving. But there's so much traffic congestion, and there are so many accidents still left that the only way to get better surface transportation is to have automated cars, automated trucks, and automated buses with all of the sensors that can help people do the driving*" [6].

1.4 Automated Highway System (AHS) Concept

New transport opportunities are closely related to appropriate infrastructure to fully exploit such the emerging capabilities. As mentioned above, another rather future oriented option is the introduction of the driverless trucks, which will reduce employment costs, improve reliability and limit the number of mistakes. However, next to new equipment, sensors, command and control centers, they will also need improved infrastructure. When concerning the complexity of such the concepts it is a rather long-term vision connected with serious research and considerable funds. One of the probable future concepts is the Automated Highway System (AHS), also called Smart Road, and it is a futuristic concept of a highly developed intelligent transportation system technology. AHS can be defined „*new relationship between vehicles and infrastructure. It refers to set of designated lanes on a limited roadway where special equipped vehicles are operated under completely automatic control*" [2]. It is one of the major international concepts in the frame of the ITS designed "*to provide the basis for, and transition to, the next major performance upgrade of the vehicle/highway system through the use of automated vehicle control technology*"[12].

The concept of "*fully automated intelligent vehicle-highway system*" is interpreted as a system that [12]:

- evolves from today's roads (beginning in selected corridors),
- provides fully automated "hands-off" operation at better levels of performance than today's roadways in terms of safety, efficiency, and operator comfort,
- allows appropriately equipped vehicles to operate in both urban and rural areas on highways that are both instrumented, and not instrumented.

The main aim of the concept is to implement driverless cars on developed roads to reduce traffic congestion by decreasing distances between cars, and at the same time allocating more cars on the roads. This is because "platooning" enables vehicles to "*operate much closer together than is possible under manual driving conditions, each lane can carry at least twice as much traffic as it can today*" [14]. Additionally, wind-tunnel tests conducted by the University of Southern California have proved that the drag force in a "platoon" or convoy can be cut in half in the case of vehicles operating at a separation of about half a vehicle length. At the same time they have proved rather significant decrease of fuel consumption (20 to 25%) with accompanying emissions reductions. However, the main idea of the study is the creation of special highways, or at least short parts of existing ones, ready to support such the purposely designed vehicles. Having many sensors onboard, such vehicles could drive automatically and safely along "smart" roads, equipped with power steering and automatic speed controls operated by onboard computer. AHS system sensors will be design to "*read passive road markings, and use radar and inter-car communications*" to organize cars in traffic themselves without the intervention of drivers. As a result, it will be possible to organize cars into convoys of eight to twenty-five cars driving without human support just about one meter apart. Parallel, air resistance will be minimized and the distance between platoons will follow typical conventional braking distance, so in the case of trouble only one convoy would be affected.

The AHS project has been started by researchers from The Ohio State University and the first automated vehicle was built in 1962, as probably the first land vehicle equipped with a computer to control steering, braking and speed. Very interesting progress was also achieved in the frame of the highway programs called the California Partners for Advanced Transit and Highways (PATH), established in 1986 by agreement between University of California's Institute of Transportation Studies and the California Department of Transportation. The mission of the program was to *"apply advanced technology to improve highway capacity and safety, and to reduce traffic congestion, air pollution, and energy consumption"*. The program focuses on areas that offer potentially quick improvements regarding the transportation systems as "the growth of population and travel demands is so rapid, that the effects of incremental solutions are likely to be absorbed by this growth by the time that they are implemented". To meet such goals, the research involved "*forty professors and about eighty graduate students at both UC Berkeley and other universities throughout*" the California. A prototype automated highway system (PATH project) was tested in San Diego, California in 1991 along Interstate 15 with technical success. In the frame of PATH concept scientists has created a block diagram of the five-layer AHS normal mode of operation control architecture, which is a result of their long-term effort. The layers are entitled as: network, link, coordination, regulation, and physical [4]. The most important concepts in the frame of link, coordination, regulation, and physical layers detailed models and corresponding control systems have been specified and tested. "*The physical layer comprises all the on-board vehicle controllers of the physical components of a vehicle*" including "*the engine and transmission, brake and steering control systems" and "lateral and longitudinal vehicle guidance and range sensors*"[4]. Their role is "to decouple the longitudinal and lateral vehicle guidance control and to approximately linearize the physical layer dynamics".

"*The regulation layer is responsible for the longitudinal and lateral guidance of the vehicle and the execution of the maneuvers ordered by the coordination layer*". It is designed to carry two longitudinal control tasks [4]:

- the first is related to vehicle-follower in a platoon and is focused on maintaining a prescribed constant spacing from the preceding vehicle;
- the second is related to a platoon leader and is focused on safely and efficiently executing a maneuver commanded by the coordination layer.

Table 1. The five layers and their main functions. Source: [4].

Layer	Function	Model
Network	Control entering traffic and route traffic flow within AHS network	Capacitated graph
Link	Compute and broadcast activity plans (i.e. the routes, maneuvers to be executed, speed, platoon, size) for each vehicle type in each section	Fluid flow model with distributed control
Coordination	Communicate and coordinate with peer and select one maneuver to be executed	Finite state machine
Regulation	Execute maneuvers such as join, spit, line change	Feedback laws based on linear models
Physical	Decouple lateral and longitudinal control	High order nonlinear differential equations

The coordination layer selects "*the activity that the vehicle should attempt or continue to execute, in order to realize its currently assigned activity plan. It communicates and coordinates its actions with its peers -the coordination layers of neighboring vehicles - and supervises and commands the regulation layer to execute or abort maneuvers*" [4]. The link layer controller "*receives commands from the network layer in the form of demands on the inlet traffic flows at the AHS entrances, and outlet flow constraints at the AHS exits, as well as desired inlet-to-outlet traffic flow split ratios, in case a vehicle can take more than one route to observe that there are far fewer such commands than the number of cars in each section reach the same destination*" [4], while traveling in that highway link. The network layer is responsible for controlling entering traffic and route traffic flow within the network of highway links that constitute the AHS, in order to optimize the capacity and average vehicle travel time of the AHS and minimize transient congestion in any of its highway links. The overall effort in the frame of PATH is clearly presenting that the progress in the field of technological support for transportation systems is quite significant. When we put together DARPA's Grand Challenges and PATH it is visible that when connected with AHS type concepts they are not futuristic ones only. This approach creates the reality of current research and the nearest future practical usage.

However, after the first promising results, investments have been moved more toward autonomous intelligent vehicles rather than building specialized infrastructure. The concept is still under progress as possible results are rather promising and for developed countries and also developing like China and India, as it is good solution for traffic congestion which is causing more and more problems. New system can help by using automated vehicles on "smart" highways for example during night or out of rush hours. In Poland there are not developed concepts to use AHS type vehicles and infrastructure, although there are many institutions capable to conduct such studies. Especially, infrastructure must be improved as soon as possible as it cannot support futuristic solutions and in the field the country will not be able to follow modern transportation trends. The concept is also presenting that the shift from infrastructure focus into independent vehicles was caused when sponsors realized high costs of not only research, but as well further implementation of new type of highways.

1.5 Underground Automated Highways (UAH)

UAH is still rather futuristic concept but it is now are under development to improve traffic flow and its safety. It was first proposed in 1963 and at present it has many followers. The study "Urban Underground Highways and Parking Facilities" provided "examples of what highway needs might be if all the users of mass transit systems were transferred to passenger cars in Los Angeles, Chicago, and Manhattan" [3]. Moreover, it provides the foundation for future studies and technical development of the underground-highway concepts. The concept of AHS has been predicted to run through decades, but at present there is similar discussion regarding implementation of UAH system concepts. Futurists and transportation experts agree "that when it comes to letting machines drive you through underground tunnels; it's probably more a question of when, rather than if" [7]. They predict that in 50 - 100 years such UAH systems will become reality perhaps in the United States or Europe, although it would be huge undertaking.

To achieve such a goal and to implement it into practice it will be necessary to experience significant advancements in the three fields [13]:

- A working AHS net as it is obligatory to welcome a similar system and technology aboveground first. This step will enable testing new smart highways and appropriate vehicles plus AHS connected with UAH will create a complete transportation system. Moreover, once all users gradually become accustomed to the technology with aboveground solutions they will be ready to go under the ground level rather easily.
- Zero-emission vehicles must be developed as an underground highway would involve a great deal of traffic whizzing through subterranean passages and it will be challenging to ventilate them without having to pump out clouds of vehicle exhaust. As a result, vehicles that produce zero emissions (fuel cells, batteries, solar power, hydrogen power or other energy-efficient methods) will be necessary.

- Improved tunnel-boring technology, as it will be necessary to dig a significant number of large tunnels. Taking into consideration the Channel Tunnel, which is 50 kilometers long and runs 60 meters underneath the English Channel; it took four years to complete it. New more effective tunnel-boring solutions are needed. Some countries have continued to develop tunneling projects, leading to a decrease in tunneling costs and an increase in efficiency ($1.50 per cubic foot, rate of six meters per hour).

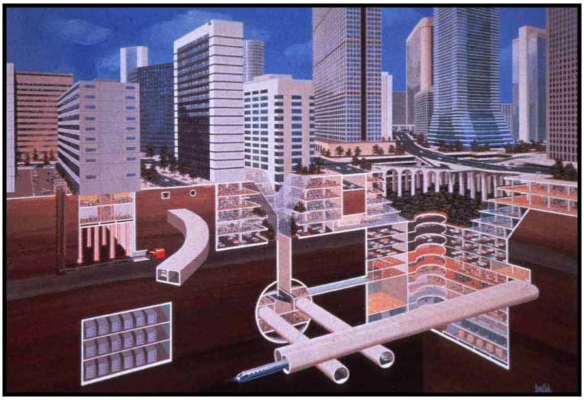


Fig. 4. Extended use of underground space. Source: [15].

However, the concept is very complex, as UAH will create additional challenges including the need to create improved earthquake protection systems. Past examples have shown that underground projects are reasonably strong and somewhat resistant so these protection measures might be put in place at reasonable cost. For example, the tragic 1995 earthquake in Japan proved that underground constructions survived it with rather minimal damages inside the Kobe city [15]. Moreover during construction, a great quantity of dirt and rock will be produced and there will a need to relocate it. Finally, at least at the beginning, people will be rather cautious before starting to use new highways, so the public must be informed ahead. According to experts if enabling technologies, mentioned above, will continue to improve, it is possible to "*forecast that construction on the first underground corridors for AH networks will begin circa 2030 in our largest and wealthiest cities, as that is about the time we are likely to have cheap and plentiful tunnel boring machines (TBM's), a limited surface-level AH network in HOV lanes in a few cities, and a significant percentage (≥ 40%) of hybrid, ultra low, or zero emission vehicles in deployment*"[11]. This is because UAH are important factor involving many benefits e.g. support for "*urban space utilization, efficient and sustainable transportation, and automating a significant fraction of the urban commute for city inhabitants*" [11].

It is necessary to emphasize that low expense and well-organized mass transit in the overcrowded cities must reverse what has been a worsening situation regarding urban transportation and to start new dimension planning, as it provides the most reasonably priced and highest density transportation option. There are also opinions, that "*the most rapidly growing complementary long term transportation system in the* *leading high density cities*" at the end of the current century will be a "*gigantic network of UAH*" [11]. Again, this is for the reason that the construction of UAH will considerably increase the safety, speed, aesthetics, and capacity of goods and human transport in cities. When combined with underground parking structures at source and destination, "*such systems promise to increase metropolitan traffic capacity and throughput by at least another order of magnitude in their presently conceivable deployment, eventually halving or thirding today's average urban commute times, and reducing surface transportation architecture, noise pollution, and visual blight through the selective takeback of some of our most valuable surface architecture currently dedicated to surface transportation*" [11].

2 Conclusions

This paper explores some concepts have been discussed regarding ideas which will be able to replace drivers and to improve the safety of logistics processes in the field of transportation. New type highways and trucks equipped with sensors to support drivers along with new automatic vehicles represent the main direction of these efforts. AHS and UAH concepts are very especially promising and are being developed. In the near future, they will probably change logistics' transportation services. Smart highways connected with driverless cars will enable better flexibility to plan and conduct shipment of goods especially during periods of less demand on road networks. Adaptation of these systems would be beneficial for all the logistics providers. Furthermore, in the case of technological support for the logistics, there is a need to take a multidimensional approach as all the elements must meet some standards to make new, long-distance transport solutions really operational. This will require that all components of the system must be developed in a continuous way. So, it is necessary to analyze these components and incorporate them with modern logistics including Polish logistics, technologies, highways etc. Such solutions will be implemented in the future, so it is necessary to prepare logistics infrastructure in the country for their needs in advance. At present the research should be focused on support for the drivers and investment in new, better vehicles, as there are strong research centers and academic institutions to provide such the support. The reason is that funds are not big enough to deal fully with new highways concepts. At the same time, it is rather necessary to merge science approach and practical improvement of highways to include even now some passive road sensors to use them in the future. The problem is quite urgent as new capabilities will be available, so modern transportation in the country must up to date to meet clients' expectations and to keep competitive with foreign companies. However, the case studies presented above, although promising, are still under development and some of them are still rather in early stages and the progress is slow. The current financial crisis will definitely influence that continuity by cutting research and implementation funds. But, new growing economic powers like India, China, will support USA and EU in their efforts, as they also are looking for new solutions and they have surplus of funds for investment in the futuristic concepts, looking for better capabilities, to improve the competitiveness of their companies and to develop new practical technologies.

References

1. Chen, Q., Özgüner, Ü., Redmill, K.: Ohio State University at the 2004 DARPA Grand Challenge: Developing a Completely Autonomous Vehicle. IEEE Intelligent Systems, Ohio State University (September/October 2004)
2. Cheon, S.: An Overview of Automated Highway Systems (AHS) and the Social and Institutional Challenges They Face. University of California Transportation Centre (2003)
3. Hoffman, G.: Urban Underground Highways and Parking Facilities. RAND Corporation, Santa Monica (1963)
4. Horowitz, R., Varaiya, P.: Control Design of an Automated Highway System. University of California, Berkeley (2000)
5. Integrated Vehicle-Based Safety Systems (IVBSS), The Transportation Research Institute, the University of Michigan, `http://www.umtri.umich.edu/divisionPage.php?pageID=249`
6. Interview with Chuck Thorpe of Navlab, NOVA Escape (2000)
7. Lamb, R.: Will we drive on underground automated highways? HowStuffWorks.com (2008)
8. Pordzik, M., Socha, L.: Control models of vehicles on three lane highway, Zeszyty Naukowe no. 41, Politechnika Śląska, Gliwice (2000)
9. Sayer, J., Leblanc, D., Bogard, S., Blankespoor, A.: Integrated Vehicle-Based Safety Systems (IVBSS) Heavy Truck Platform Field Operational Test. Data Analysis Plan, the University of Michigan Transportation Research Institute (UMTRI), Michigan (2009)
10. Shladover, S.E.: Automatic Vehicle Control Developments in the PATH Program. University of California Berkeley, Transportation Library, Berkeley (1991)
11. Smart, J.: Underground Automated Highways (UAH) for High-Density Cities. Acceleration Studies Foundation (2005)
12. Ünsal, C.: Intelligent Navigation of Autonomous Vehicles in an Automated Highway System: Learning Methods and Interacting Vehicles Approach, Virginia Polytechnic Institute and State University, Blacksburg (1997)
13. Underground Automated Highways (UAH) for High-Density Cities: A 2030-2060 Forecast, The Coming Convergence of Three Technologies: A Post-2030 Developmental Attractor, Acceleration Studies Foundation (2005)
14. Vehicle Platooning and Automated Highways, California PATH–Partners for Advanced Transit and Highways, `http://www.path.berkeley.edu/PATH/Publications/VPlatooning.pdf`
15. Why go underground?, International Tunneling Association – Association Internationale des Travaux en Souterrain (2002)
16. Zeybek, H.: Intermodal Freight Transport and Logistics Research. In: European Union and Turkey, Ankara (2010)
17. The Research and Innovative Technology Administration (RITA), `http://www.its.dot.gov/ivbss/index.htm`

“Defence-in-Depth” Strategy in Transport Risk Management

Andrzej Szymanek

Technical University of Radom,
Malczewskiego 29, 26-600 Radom, Poland
a.szymanek@pr.radom.pl

Abstract. Safety management is a kind of system management, that is management by purposes. Taking “defence-in-depth” strategy, DDS – there can be defined four main aims and four method groups of risk management in transport: 1. minimizing transport accidents risk; 2. minimizing number of undesirable transport events (incidents, conflicts, collisions, accidents). Above purposes relate stages of safety management in transport. At each level of management should be elaborated methods, procedures and technologies of minimizing transport accidents risk. According to DDS any management system of transport safety should have a structure of multilevel chain protections which supervise main transport processes. About those problems in the paper.

Keywords: Defence, safety, barriers, transport, risk, management.

1 Introduction

Managing transport safety is de facto risk management, that is managing loss avoiding, which can occur at every level of functioning and in every place of concrete transport system.

Safety management in transport is a managing by goals which in every transport system can be defined, taking as a starting point circumstances of loss arising. The main goals of transport safety management are:

1st goal: minimizing risk of pre-accidents events (incidents, conflicts, collisions, accidents);

2nd goal: minimizing risk of transport accidents;

3rd goal: minimizing transport accidents effects;

4th goal: minimizing effects of transport catastrophes.

Realizing above goals require applying procedures and technologies which create multilevel protection system of moving processes, load processes and processes of traffic control. This is an approach consistent with “defence in depth” strategy. It can assure effective safety management, that is such managing where “risk level” is relatively constant and probability of accident is inversely proportional to “consequences”; it relates also transport.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 51–58, 2010.

2 Accident "Anatomy" – Premises for Defence-in-Depth Strategy

Observations of accidents shows that sudden changes of normal situation happens rarely, and most of accidents are preceded by obvious disturbances or tolerated danger. From among older hypothesizes of accident arising "theory of chain of branched accidents" by H. Hepburn suppose that, the condition of accident happen is convergence two (or more) series of events. Other researchers similarly reasoned: Winsemius (1958) interpreted accident as a "result of convergence few accident reasons". While Andriesen talked about "multi-causality" and "co-determination". In "domino effect" conception by H.W. Heinrich – accidents are preceded by events which create "reason chain" made of five links, [1]. In anthropo- and social-technical systems those links are interpreted as follows: 1 – operator reliability; 2 – operator fault; 3 – dangerous behavior; 4 – accident; 5 – injury, damage, [2]. Such schema can be related also to accidents in transport systems. The accident can be prevented by stopping "domino effect" – before/in point 3. By L. F. Bird and E. Adams the most important chain links of accident are those which relate with management, [2].

Each accident is the last link of sequence (chain) of events and actions, Fig.1.

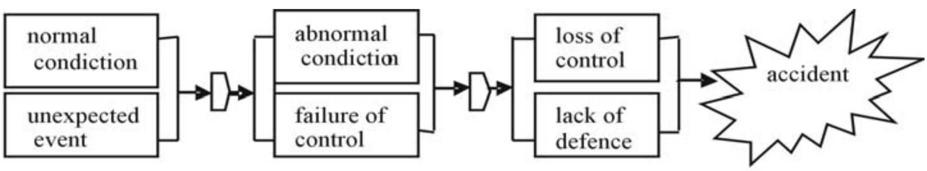


Fig. 1. The " anatomy of an accident" [3]

The hypothesis research of accidents arising and process of pre-accidents events is a domain of many safety models:

- hidden conditions models, e.g. "iceberg" [4];
- convergence models, e.g. blunt and sharp end model [5];
- process models, e.g. OARU model [6];
- phase models, especially Haddon's matrix [7];
- pathological models, especially "Swiss cheese model" [8].

Many types of accidents are an effect of coincidence different factors. The basic mechanism of preventing the "coincidence states" is steering – meant as avoiding of "coincidence states". The goal and rule of steering is then identifying potential states of coincidence and time-and-space separation collision factors.

In transport systems the point is that the first of all is steering of: 1. collision stream of vehicles; 2. disturbed each other information streams.

Every defence strategy which contains adequate answer to "sequential mechanism" of accidents arising is proper; it is compatible with thesis: what is the accident "anatomy" – that is such method of accident preventing. Such strategy is "defence-in-depth" strategy where the essence is construction of barrier chain which reduce accident risk and limit accident effects.

3 "Defence-in-Depth" Strategy and Its Usefulness in Transport Risk Management

Defence in depth this is: "A series of levels of defence (inherent features, equipment and procedures) aimed at preventing accidents and ensuring appropriate protection in the event that prevention fails" [9].

"Defence-in-depth" strategy is stopping accident chain by chain of physical protections, procedural and organization protections, where infringement one of them is detected at local level of system management. The basic protection designed according to this strategy is [10]:

1. "process supply" (safety technologies, safety procedures) – their part is providing process in normal situation;
2. safety systems – their part is realizing protection action in case of process disturbances;
3. safety barriers – their part is stopping (slowdown) accident chain;
4. safety zones – their part is limiting accident effects.

Defence in depth strategy is based on few rules which explain "laminated" nature of real safety systems; these are:

1. comprehensiveness – protection has to cover all system aspects (organization, functional, technical);
2. coordination – system protection has to be coordinated, it means that protection ways are proper to dangers detected in system and proper to warnings about those dangers;
3. dynamism – system protection has to be dynamic, so it should be identified by: reaction ability, danger scale, action plan;
4. adequacy – protection has to be adequate to dangers; it relates means of safety, system of danger detection, procedures of reactions;
5. completeness – each element of system has defence priority and several defence barriers; the minimum number of defence barriers should be proportional to risk size.

Defence-in-depth strategy is a strategy of "common sense" which recommends technology and safety procedures redundancy; its basis is organization model with potential danger analysis and linear defence estimation.

"Defence-in-depth" conception connects with many other safety fundamentals, like [11]: fail safe, inherent safety; redundancy and diversity of system; sobustness of system, which connects with fault tolerance of system function; resilience of system; Safety Integrity Level, SIL.

"Defence-in-depth" strategy requires creating many barriers, which task is elimination and reduction activating in turn dangers, what shows blunt- sharp-end model. The main issue is avoiding danger activating distant in time (closer to "blunt end") and stopping local dangers ("now and here") which are at "sharp end". Each management level has an influence on accident risk: that influence is far away in time ("blunt end") and close in time ("sharp end"). "Depth-defence" strategy order to avoid faults and using defence means – as soon as possible from blunt end.

Fig. 2. shows levels of Road transport safety management:

1: reduction of social acceptation for high level of Road accidents risk (education in range of safety culture);
2: transport policy and programs of transport safety improvement (government activities);
3: effective system of transport management – level of local government and Road administration;
4: work safety management of transport operators (carriers);
5: minimization of accidents risk by:

- anticipation of local and temporary danger factors;
- safety technology of active car;
- intelligent transport systems in traffic management;

6: decrease of road accidents effects:

- techniques of cars passive safety,
- automatic systems of accident alarm,
- road rescue systems,
- other.

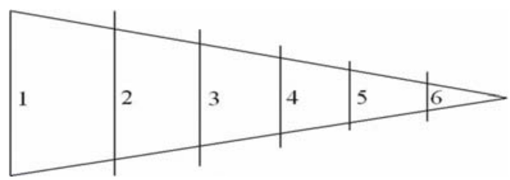


Fig. 2. Blunt-sharp-end model and levels of road transport safety management (self work)

4 Safety Barriers and Risk Management in Transport

Safety barriers play a big role in understanding and accidents prevention. ISO17776 norm defines safety barrier as a "mean which reduces possibility of getting real potentially dangerous losses and calm down its consequences" [12]. IEC 61511-1 norm defines defence layer as "any independent mechanism which reduces risk by control, prevention and weakness" [13]. By other definition barrier is: "a measure which reduces the probability of realizing a hazards potential for harm and of reducing its consequences. Barriers may be physical (materials, protective devices, shields, segregation etc.) or non-physical (procedures, inspections, training, drills)" [14]. Similar interpretation is given to safety function: "a safety function is defined as a technical, organizational or combined function, which can reduce the probability and/or consequences of accidents and other unwanted events in a system" [11].

The most often used safety barrier classification is the following [4]:

- material barriers, physical: simple device and constructions, which stop, limit, disturb dangers;
- functional barriers: counteract dangers by use the values: times of delays, distances, synchronizations, which influence on safety system functioning;

- symbolic barriers: they use different forms and information (codes, ciphers, instructions, procedures, reports) to prevent dangers in system;
- material barriers: rules, law regulations, supervision.

Basic parameters of safety barriers quality are: effectiveness, reliability, costs. Barrier effectiveness, that is probability that barrier achieves proper safety functions depend as barrier is manage during its life cycle. In methodology of estimation integrated risk of accidents at industrial areas the most important functions of safety barriers are [15]:

- risk identification – 23%;
- distribution of roles, responsibilities for barrier – 13%
- monitoring, feedback, learning and change management - 10%
- provision of competence and safety suitability – 14%
- management of procedures, rules and goals – 10%.

Among used models in safety management there are barrier models. Especially it's all about "Swiss cheese" model by J. Reason [8] and AEB model (Accident Evolution and Barrier Function) [16]. In the first model there is a thesis, that accident can happen as a result of overlapping organization faults, dangerous behaviors of participants, hidden dangerous conditions, mistaken supervision activities and leaky defence barriers (that is "holes" in Swiss cheese). That model was used in air accidents researches.

In AEB model there is a barriers chain which prevents accident by stopping human faults, organizational and technical faults. Such model was used in management of railway and sea management [3], [4]. In AEB are required: potential danger identification, establishment of defence goals and next determining which barriers should be used in purpose of accident prevention. Main message of barrier analysis is a motto: "keep danger far away from goal".

In "defence-in-depth" strategy for informatique systems safety management "barrier" is defined as an element which goal is protection of informatique system against dangers. With barriers integrated program of their monitoring has to cooperate. With barriers has to cooperate integrated program of their monitoring. However "defence line" is a barrier team which after breaking generates an event, which danger level depends on amount of intacted barriers left among danger and protected resources. One barrier can protect many resources, but not necessary in the same way. As a result defence lines should be analyzed separately for each resource and separately for each danger (or group of dangers) [17].

5 Four Goals and Three Levels of Risk Management in Transport Systems. Conception of 4G-3S Model

Mentioned before four main goals of safety management in transport can be realized at three different "levels" of transport system. 3-STR conception of "three levels of transport risk" bases on three different definitions of general system:

1. structural definition – relates to inner system construction;
2. functional definition – relates to system functioning, which identifies by processes which change system characteristics;
3. simulation-model definition – relates system in defined work conditions.

According to 3-STR conception risk analyses should be provided at three stages:

1. transport system structure;
2. working processes in transport system;
3. transport system "behaviors", that is de facto – stage of negative transport effects (ENTs).

Three stages of transport risk are following:

1 stage: structural risk SR – the effects of all undesirable interactions between elements of transport system structure; factors of structural risk, that is among others: human factor, means of transport, transport infrastructure, "wrong matching mean of transport to human-operator", "wrong reading the elements of transport infrastructure"; effects at "contact point": mean of transport – infrastructure of transport, other;

2 stage: functional risk FR generated by undesirable changes in traffic processes, load processes, processes of traffic control in transport system. Especially - by interpretation of process approach to transport system functional risks can be identified related with:

2.1. creating process of transport infrastructure: a. infrastructure planning, b: infrastructure realization, c: infrastructure exploitation;

2.2. process of transport services realization;

2.3. process of creating transport policy.

3 stage: system risks (behavioral) BR which generate "finish" effects, that is transport accidents, traffic congestions, natural environment degradation.

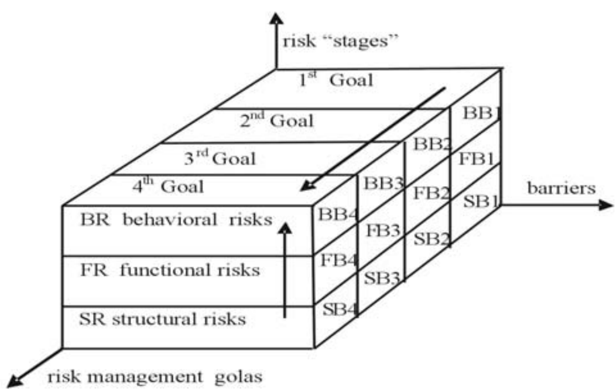


Fig. 3. General conception of 4D-3S model of transport risks management, (self study)

The most often there are risks undesirable transport events, which define transport safety problem. It is all about transport accident risk, death risk in transport accident, risk of injury in transport accident, risk of transport catastrophe. Comparing 3-STR conception with 4G conception of four goals of risk management there can be proposed 4G-3S model, that is model of "four goals and three stages of transport risks management", Fig. 3.

Here: SB – structural barriers , FB – functional barriers ; BB – behavioral barriers.

According to "depth-defence" strategy safety barriers have to exist at levels: system structure, working processes of system, transport system behaviors, and they have to be dedicated every time for protection realization four main goals of transport safety system.

6 "Defence-in-Depth" Strategy and Road Incident Management

Road incidents management is a planning and realization tasks of reactions on random road events, which can: 1. decrease traffic safety; 2. increase traffic congestion. Fast and reliable detection of road incidents and effective reaction on their appear – these are goals of automatic detection road incidents system

Elements of "depth-defence" strategy can be used to road incident management, Fig. 4. Safety barriers are detectors stations d_i , d_{i+1} occupacy of road element. Each pair of detectors is a defence line before incidents which can appear at part between those detectors. Detecting incident move two defence processes:

- for vehicles which were before incident place – that is process soft prevention about incident happen at section [d_{i-1}, d_i];
- for vehicles which are at road section, where was an – this is process of incident management.

Next defence lines are procedures of incident management, that is procedures and activities which remove incident results, among others traffic congestion elimination.

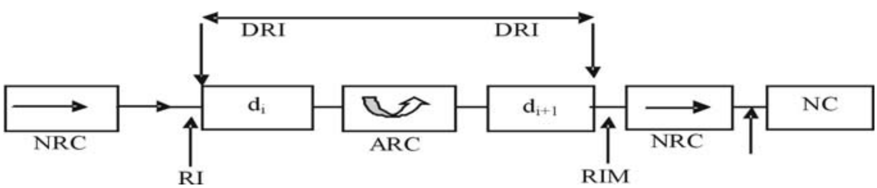


Fig. 4. Depth-defence strategy and automatic road incident detection (self study)

NRC – normal road condition; ARC – abnormal road condition, congestion;

d_i , d_{i+1} – neighboring detectors stations of road section; RI – road incident; DRI – automatic incident detection – first barrier; RIM – road incident management – second barrier (intervention actions or rescue, information for drivers, temporal change of traffic organization).

7 Conclusions

Depth-defence strategy is one of important rules of safety. According to it – each action system should be "equipped" with a chain of multiple physical , technical, procedural and organization protections which improve safety. Encroach any of safety system barriers is detecting at local level. Depth-defence strategy orders to use defence means – as earliest – starting from high levels of safety system management, so such levels which influence at level losses risk in system seems to be far away in time. It has a relation with safety culture.

"Depth-defence" strategy should be wider used in transport safety management. It seems to be obvious using for example multiple independent barriers in software and intelligent constructions of transport systems. The result will be significant improve of traffic safety and transport structure efficiency.

References

1. Heinrich, H.W.: Industrial accidents prevention: A scientific approach. McGraw-Hill Book Co., New York (1936)
2. Dong, Z., Niansheng, X.: Accident causation and prevention theories. In: Qingxuan, Z., et al. (eds.) Progres in Safety Science and Technology. Proceedings of the 98 International Symposium on Safety Science and Technology, pp. 61–66. Science Press, Beijing (1998)
3. Hollnagel, E.: Accident analysis and barrier functions. IFE (N). Version 1.0, 99/09/30 (February 1999)
4. Hollnagel, E.: Barier Analysis and Accident Prevention. Human - Technology Integration Colloquium Series. Air Forses Research Laboratory, Finland (2002)
5. Cook, R.I., Woods, D.: A Tale of Two Stories: Contrasting Views of Patient Safety. In: Report from a Workshop on Assembling the Scientific Basis for Progress on Patient Safety. National Patient Safety Foundation at the AMA, pp. 13–14(1998)
6. Kjellen, U.: Prevention of accidents througth experience feedback. Taylor & Francis, London (2000)
7. Runyan, C.W.: Using the Haddon Matrix: introducing the third dimension. Injury Prevention 4, 302–307 (1998)
8. Reason, J.: Human Error. Cambridge University Press, Cambridge (1990)
9. Reason, J.: Managing the Risk of Organizational Accident. Aldershot, Ashgate (1997)
10. Pasman, H.J.: Safety will save costs, but how much ist he question! In: Qingxuan, Z., et al. (eds.) Progress in Safety Science and Technology. Proceedings of the 1998 International Symposium on Safety Science and Technology, Beijing, New York, pp. 371–372 (1998)
11. NORSOK Standard, Design Principles /Technical Safety, Norwegian Technology Standards Institution, Oslo, S-DP-001, Rev.1 (December 1994)
12. ISO 17776 Petroleum and Natural Gas Industries – Offshore Production Installations-Guidelines on Tools and Techniques for Hazard identification and Risk Assessment (2000)
13. Functional Safety: Safety Instrumented Systems for the Process Industry Sector, Part 1: Framework, Definitions, Systems, Hardware and Software, Requirements (2003)
14. NORSOK Standard, Technical Safety, Norwegian Technology Standards Institution, Oslo, S-001, Rev.3 (January 2000)
15. Duijm, N.J., et al.: Evaluating and Managing Safety Barriers in Major Hazard Plants. In: PSAM 7– ESREL 2004. Springer, Berlin (2004)
16. Svenson, O.: The accident evolution and barrier function (AEB) model applied to incident analysis in the processing industries. Risk Analysis 11(3), 499–507 (1991)
17. In Depth Defence applied to Information Systems. Version 1.1 – (July 19, 2004); This document was produced by the DCSSI Advisory Office (SGDN / DCSSI / SDO / BCS)

The Problems of Railway Infrastructure Modernisation and Their Influence on the Modernisation of Regional Routes

Andrzej Białoń[1,2], Paweł Gradowski[2], and Marta Gryglas[2]

[1] Faculty of Transport, Silesian University of Technology,
Krasińskiego 8, 40-019 Katowice, Poland
[2] Railway Institute, Railway Traffic Control and Telecom Unit,
Chłopickiego 50, 04-275 Warsaw, Poland
{abialon,pgradowski,mgryglas}@ikolej.pl

Abstract. The article presents the assets of PKP PLK S.A. company as well as their infrastructural condition, which is undergoing decapitalisation. Current modernisation projects require feasibility studies. Those studies include issues related to modernisation, which in turn influence the efficiency of these projects. These issues have been discussed briefly. The problems of modernising low-traffic routes have also been presented.

Keywords: PKP PLK S.A., investment efficiency, feasibility study, modernisation issues, modernisation of low-traffic routes.

1 Introduction

PKP Polskie Linie Kolejowe (PLK) S.A. is one of the biggest administrators of railway infrastructure in Poland. Like any enterprise operating in the market, the company has certain assets. PKP PLK S.A. assets consist of fixed assets (86,4%), that include: land, buildings and facilities, civil engineering facilities, technical equipment and machines, tangible assets in progress, with advanced payments; remaining assets, as well as current assets (13.6%) include: reserves, receivables for supplies and services, tax, custom, insurance and other provisions, any other receivables, short-term financial assets, prepayments and accruals.

1.1 Restructuring and Rationalising the Assets

The process of bringing PKP S.A. assets required for railway route management to PKP PLK S.A. is gradual according to the provisions of the Act of September 8, 2000 concerning commercialisation, restructuring and privatisation of "Polish State Railway" state company. At the same time, the railway infrastructure is being adjusted for statutory activities. Any redundant assets are liquidated or transferred to local governments according to law.

Home investments produced a huge interest among entrepreneurs in reactivating railway traffic on routes that were already cancelled or decided to get cancelled. Following

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 59–66, 2010.

an economic analysis to show the profitability of this action, the Board of PKP PLK S.A. have agreed to regenerate more stretches or railway routes in recent years.

PKP PLK S.A., who is the main administrator of the railway infrastructure, assigned a basic goal to the remaining routes, i.e. to maintain the railway network in a condition allowing for safe railway traffic, maintain the quality of services and develop the infrastructure.

2 Railway Routes Operation

The distance of operated routes year by year (in 2008) has grown a little, which is a positive market trend. But network-wise, the most important goal for the administrator of the infrastructure is to provide the passenger and cargo carriers, as well as their passengers and consigners with a high-standard offer. That is why improving the technical condition of the infrastructure is a priority for PKP PLK S.A. Due to this, the maintenance and repair works conducted by the company are concentrated mainly on removing any speed limitations.

Table 1. The structure of scheduled speeds

The structure of scheduled speeds		
Speed ranges	Rail length (km)	Total rail length share
V >160 km/h	1.493	5%
120 < V < 160 km/h	4.011	15%
80 < V < 120 km/h	10.482	38%
40 < V < 80 km/h	9.259	33%
V < 40 km/h	2.534	9%
Total	27.779	100%

Unfortunately, the lack of finances to cover infrastructural repairs means that its condition is systematically deteriorating. When introducing new schedules, the length of rail sections with lowered scheduled speed exceeds by approximately 1,000 km the length of rail sections, on which this speed was increased. The scheduled speed structure in the 2008/09 schedule is presented in the Table 1.

Decreasing the scheduled speed in new schedules has not contributed to lowering the number of speed limitations nor to the length of reduced-speed sections. A comparison of reduced-speed rail sections in consecutive years is shown in Fig. 1.

In order to improve the technical condition of the infrastructure, increased repair and modernising works are required.

That is why the main aim of the modernisation is to improve the rail route parameters to meet the new standards and the requirements of the AGC international agreements (concerning main international railway routes) and AGTC (concerning the main railway routes in combined transportation).

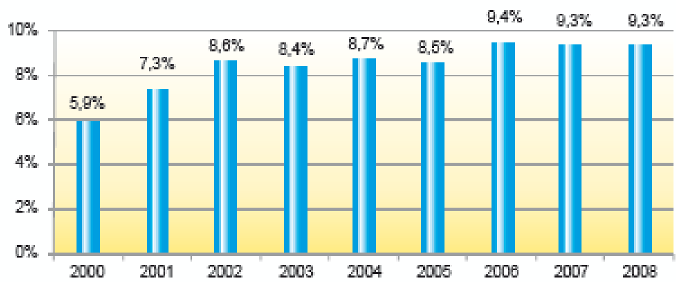


Fig. 1. The length of reduced-speed rail sections

The current and future modernisations are expected to increase the comfort and safety of rail transportation as well as decrease the maintenance cost. Modernisation will also contribute to environmental safety - all projects include provisions to protect the environment from noise, vibrations, soil and water pollution. The speed will be greatly increased by up to 160 km/h for passenger trains and 120 km/h for freight trains, and by using the Central Railway Line - up to 200 km/h. Modernisation of the railway network can also result in decreased traffic density on public roads.

3 Perspectives for Operated Railway Routes

The condition of Polish railway infrastructure has a significant influence on the development of Polish enterprise, new investments, the competitiveness of railway operators against road transportation and attracting individual customers. The process of modernising Polish infrastructure is insufficient. The estimated funds provided for this task cover only 20% of the required amount.

The necessary annual costs of infrastructure renovation, assuming a 10-year period needed to overcome the repair backlog while maintaining annual rail and turnout replacements caused by wear and tear, are 3.5 billion PLN to overcome the backlog and 1.2 billion PLN for current works. 2.2 billion PLN should cover the rail and turnout maintenance as well as diagnostics, while 250 million PLN should cover the costs of replacing heat-treated rails. This means that the annual costs for the next 10 years should amount to 7.150 billion PLN. Obtaining such an amount does not seem to be viable.

It is, thus, necessary to design a new concept of maintaining and modernising the infrastructure and, above all, of rational division of funds among maintenance and investments, including EU funds available in 2007-2013. In 2006, the European Union provided Poland approximately 25 billion euro for infrastructural investments during the current EU budget term. Increased funds, although still insufficient, have influenced the increased investments in PKP PLK S.A. as well as the development of products and services in this sector. Returning to the normal operating conditions on main routes is, however, still the most important element in the next 10 years. Investment policies should be changed to favour these works that result in the visible shortening of travel time for as many passengers as possible, getting rid of bottlenecks in the network and increasing investment in traffic automation. Apart from increasing infrastructural financing, cost rationalisation at PKP PLK S.A. is also necessary.

Unfortunately, the modernisation works as well as new investments in railway routes are progressing too slowly, making the overall improvement in infrastructure very small.

3.1 Railway Problems

As is the case in most European countries, the Polish railway network was formed when the railway transportation share in the overall transportation was the greatest. Any problems with supporting an extensive railway network were mostly caused by the decrease in transportation. In some cases, the question of the advisability of maintaining certain sections can be posed, but their cancellation or abandoning their renovation would still be a mistake.

The Feasibility Studies conducted so far have shown that there is a gap in planning the investments in the railway infrastructure. This results in a lack of an overall concept concerning the biggest railway junctions, which would answer the challenges and fit the plans of the city and agglomeration development.

Charges paid by carriers for accessing the railway infrastructure, while still insufficient to fully cover the cost of operation and maintenance of this infrastructure, also serve as an important economic factor to be considered. This poses a threat of unbalancing the cost of repair, maintenance and operation of the existing infrastructure with the national treasury contributions and proceeds from customers.

Abandoning the turnout, rail and ballast replacements as well as engineering maintenance along the routes, which should be performed periodically, as well as other repairs will cause the route parameters, especially the speed limits, for keeping transportation safe to deteriorate . As a result, the speed limit on many routes is less than that scheduled, decreasing the comfort.

The rolling stock should also be replaced in parallel to the modernisation, as the current stock is and will be insufficient to meet the prospected needs. So far, any purchases of passenger vehicles were made on a unit-by-unit basis (e.g. several diesel rail buses bought by Marshall Offices, passenger cars, a dozen electric locomotives for passenger trains) or were short EZT series (e.g. to operate city-to-city routes and push-pull trains). On a global scale, these purchases did not cover the losses in rolling stock, improved the offer in a limited and selective way, and introduced a variety of models, which is harmful for maintenance and repair purposes. There have been incomprehensible situations in the process, such as the purchase of modern push-pull trains or traction vehicles to be operated remotely, which means that these modern trains are pulled 1960s locomotives, while their own control rooms remain defunct. The manufacturers of passenger trains have repeatedly proposed more effective ways to purchase new rolling stock, such as long-term planning, series extensions and option purchasing. These possibilities were, however, never used.

4 Modernisation Opportunities

By analysing the network size, we can determine the potential market of low-traffic routes, where a simplified traffic control system can be implemented in the future. Route no. 27 between Nasielk and Sierpc, which is a part of the Nasielsk - Torun

Wschodni route, can be given as an example of a low-traffic route fit for further analysis. Even though there are only 4 stations along this route, this section is a good solution, as the Nasielsk station would be a combination (as a Local Control Centre - LCC) of a main line (E 65 Warsaw-Gdansk) with a low-traffic route operated by the LCC.

4.1 Section Characteristics

The Nasielsk - Sierpc section located along route no. 27 Naielsk - Torun Wschodni is 87.489 km long. It is a single-rail, non-electrified route with a speed limit of 100 km/h. Due to the degradation of rail infrastructure in the Nasielsk - Sieprc section, the trains have to slow down to 60 km/h.

There are 4 stations along this section: Nasielsk, Plonsk, Raciaz i Sierpc, as well as 10 passenger stops: Cieksyn, Wkra, Dalanowek, Arcelin, Baboszewo, Kaczorowo, Koziebrody, Zawidz Koscielny, Zawidz and Mieszaki.

The stations contain key mechanical equipment (2 stations), i.e. E and PB-type transmitters. Individual sections are equipped with automatic block systems with mechanical blocks. There are 5 station (cat. A) and 82 route (cat. D) railway crossings along the route.

4.2 Modernisation Variants

There are three variants for installing a simplified CTC system for low-traffic routes in this theoretically considered route:

- Variant I - implementing train control devices for main lines in this route, while modernising the infrastructure;
- Variant II - implementing simplified train control systems for low-traffic routes;
- Variant III - implementing simplified train control systems for low-traffic routes while modernising the infrastructure to overcome low scheduled speeds (below 40 km/h).

4.3 The Results of Variant Analysis

Calculations were made for the above variants to check the influence of individual works on the analysed indices that will vary as the works progress and will consist of the combination of different works. In case of variant I, it was assumed that the “full” CTC devices and category A and B crossings will be implemented in the stations, while in the case of variant B, simplified CTC devices and category A and C crossings are planned. The Table 2 presents a comparison of the projected expenses.

When analysing the values in Table 2, we can assume that:

- In case of the modernised railway infrastructure owned by PKP PLK S.A. on low-traffic routes, as far as financial matters are concerned, installing “full” (widely used) CTC devices are advisable. Considering the condition of current rail infrastructure (significant decapitalisation), this variant is not very viable, even though it is possible. Installation of simplified CTC devices (positive financial indices despite route decapitalisation) throughout the whole term of 30 years will result in future negative indices caused by necessary rail repairs (see Table 2).

Table 2. Investment expenses on the modernisation of the example low-traffic route

	Variant I (in thousands PLN)	Variant II (in thousands PLN)	Variant III (in thousands PLN)
Rails[1]		0,00	
Renovation of 100% rail length	153,729.00		60,366.60
Renovation of 75% rail length			47,243.70
Renovation of 50% rail length			34,120.80
Renovation of 25% rail length			20,997.90
CTC device installation	39,263.40	33,313.40	33,313.40
Telecommunication	40,439.53	40,439.53	40,439.53
Railway crossings (installation of automatic crossing signalling devices)[2]			
All crossings along the route	76,634.00	68,680.00	68,680.00
Every 2 km	39,939.00	35,962.00	35,962.00
Every 3 km	27,409.00	24,790.00	24,790.00
Every 4 km	22,039.00	20,002.00	20,002.00

- Low financial effectiveness of this investment means that it can only be justified by overall economic or social benefits. Such assessments should be made in economic analyses of individual projects for low-traffic routes.
- It is hard to determine the most beneficial investment variant based on the present analysis.

Due to the current under-financing of railway infrastructure, the main purpose of charging for the access to PKP PLK S.A. network is to generate revenue and maintain regional routes, making the charges for different types of trains unique in Europe (very low prices for passenger trains, very high prices for freight trains). It has to be considered that the future state financing of railway route operation as well as regional routes will approach European standards, resulting in reduced charges for freight trains in order to relieve road traffic and in possible increased charges for passenger transportation.

Due to the above factors, it has been checked how the indices will behave in case of a 15% increase in the listed charges throughout the whole period, while maintaining the current number of trains and increasing their number to 15 per day as presented in Table 3.

Based on the analyses, it has been found that it is still hard to determine the most beneficial investment variant after the charges are raised along with increased traffic, but it can be seen that any changes in these parameters influence the final values of the analysed indices. Variant III, which gets rid of bottlenecks and increases the speeds, is the one most affected. Due to the above factors, the decision concerning the installation of simplified CTC devices in low-traffic routes will have to be preceded by a detailed analysis of given investment circumstances.

---

[1] Repairs of certain rail sections is possible in Variant III to get rid of “bottlenecks”. That is why various rail portions are considered.

[2] Due to a large number of crossings along the example route, various numbers of crossing signalling devices are considered.

Table 3. A comparison of analysis results - current traffic

	Variant I			Variant II			Variant III		
	FNPV/ C[3]	**FIRR/C**[4]	**B/C**[5]	**FNPV/ C**	**FIRR/C**	**B/C**	**FNPV/ C**	**FIRR/ C**	**B/C**
1*	25,6	12.0%	1.8	11,3	9.2%	1.5	20,6	11.7%	1.8
2**									
min	-235,8	-10.41%	0.23	-114,4	-6.4%	0.3	-162,0	-8.1%	0.3
max	179	6.1%	1.10	11,3	9.2%	1.5	-2,4	5.5%	1.0
3***									
min	-400,5	∞	0.2	-114,4	-6.4%	0.3	-230,8	-13.6%	0.2
max	-139,2	-5.5%	0.4	11,7	9.2%	1.5	-2,8	5.5%	1.1

* – Installation of CTC devices[6]
** – Installation of CTC devices, telecommunication, crossings[7]
*** –Rail works, installation of CTC devices, telecommunication, crossings[8]

Table 4. A comparison of analysis results - increased charges and traffic

	Variant I			Variant II			Variant III		
	FNPV/ C	**FIRR/ C**	**B/C**	**FNPV/ C**	**FIRR/C**	**B/C**	**FNPV/ C**	**FIRR/ C**	**B/C**
1*	48,411	15.77%	2.51	13,492	9.70%	1.59	41,010	15.73%	2.56
2**									
min	-212,9	-5.6%	0.3	-112,2	-5.5%	0.3	-141,6	-3.9%	0.4
max	23,0	9.3%	1.5	13,5	9.7%	1.6	18,0	9.0%	1.1
3***									
min	-377,6	-12.3%	0.2	-112,2	-5.5%	0.3	-210,4	-6.8%	0.3
max	-116,3	-2.0%	0.5	13,5	9.7%	1.6	17,6	8.9%	1.5

* – Installation of CTC devices[6]
** – Installation of CTC devices, telecommunication, crossings[7]
*** – Rail works, installation of CTC devices, telecommunication, crossings[8]

5 Conclusions

As has been shown above, the modernisation of low-traffic routes may seem economically unadvisable while keeping the access fees and traffic intensity on the same level.

---

[3] FNPV/C - financial net prezent value
[4] FIRR/C - financial internal rate of return
[5] B/C – benefit/cost ratio
[6] Installation of simplified CTC devices with already modernised rail infrastructure.
[7] Values for different combination of works, including the installation of simplified CTC devices with already modernised rail infrastructure.
[8] Values for different combination of works, including the installation of simplified CTC devices.

When planning the modernisation of a given route, one has to conduct detailed economic analyses of several variants every time. When completing such analyses, one has to include potential changes to access fees as well as state or local funding of railway infrastructure.

A significant profit from modernisation is possible for many low-traffic routes, provided the correct variant is chosen and traffic as well as the access fees are increased.

References

1. Updated national spatial development plan, Warsaw, Government Centre for Strategic Studies (October 2005)
2. Annual Report 2008 PKP Polskie Linie Kolejowe S.A., Warsaw, PKP Polskie Linie Kolejowa S.A (2009)
3. Poland 2030, Warsaw, Chancellery of Prime Minister of Poland, Board of Strategic Advisors to the Prime Minister (July 2009)
4. Master Plan for railway transportation in Poland until 2030, Warsaw, Ministry of Infrastructure (August 2009)
5. The White Book, Map of Polish railway issues, Warsaw-Cracow, Railway Business Forum, Transportation Committee, Polish Academy of Sciences (December 2009)
6. A list of unit prices for accessing railway infrastructure managed by PKP PLK S.A. valid from December 14, signed by the Head of the Office for Railway Transportation, Decision no. TRM-9110-09/08 of April 10 (2008)
7. Requirements for the simplified centralised traffic control system for low-traffic routes, Stage 2, study CNTK 4292/10, Warsaw (September 2008)
8. Guidelines for the economic and financial analysis of ERTMS implementation in PKP. Polish Economic Society, Katowice office, Katowice (June 2004)
9. Mikulski, J., Młyńczak, J.: Wykorzystanie systemu monitoringu GPS do oceny parametrów energetycznych lokomotyw spalinowych. Przegląd Elektrotechniczny 9, 268–272 (2009)
10. Młyńczak, J.: Układ napęd zwrotnicowy - rozjazd. Problemy diagnostyczne. Infrastruktura Transportu 1, 29–31 (2010)
11. Luft, M., Szychta, E., Szychta, L.: Method of designing ZVS boost converter. In: Proceedings of the 13th International Power Electronic and Motion Control Conference, Poznań, pp. 478–482 (2008)

The Changeable Block Distance System Analysis

Andrzej Lewiński[1] and Andrzej Toruń[2]

[1] Kazimierz Pułaski Technical University of Radom,
Faculty of Transport and Electrical Engineering,
Malczewskiego 29, 26-600 Radom, Poland
a.lewinski@pr.radom.pl
[2] Railway Institute, Railway Traffic Control and Telecom Unit,
Chłopickiego 50, 04-275 Warsaw, Poland
atorun@ikolej.pl

Abstract. The paper treats about efficiency analysis in Changeable Block Distance (CBD) System connected with wireless positioning and control of train. The analysis is based on modeling of typical ERTMS line and comparison with actual and future traffic. The calculations are related to assumed parameters of railway traffic corresponding to real time – table of distance Psary – Góra Włodowska from CMK line equipped in classic, ETCS Level 1 and ETCS with CBD systems.

Keywords: ERTMS, Changeable Block System, moving block section, efficiency analysis.

1 Introduction

The idea of Changeable Block Distance (CBD) presented in the paper is based on the following assumptions;

- The analyzed line is based on the central railway line in Poland (CMK) equipped with four state block system and typical railway control devices assuring the 160 km/h speed. The proposed modernization assumes the ETCS Level 1 system (without GSM-R) used virtual block distance, where information about signaling states assuring the 200 km/h speed was transsmited by balises.
- The hypothetical line assumed for CBD calculations corresponds to CMK structure but with ETCS Level 2 (GSM-R), the balises placed in rails are used for zero of train odometer only and deck ETCS computer transmits the telegram about train position and speed via GSM-R to RBC centre (the precise position is calculated continuously as a distance from last balise).
- The passenger trains with maximal length 400 m (average 300 m) may travel with 160 km/h speed, the truck trains with maximal length 700 m (average 600 m) may travel with 120 km/h speed.
- With ETCS Level 1 the traditional, constant block distance is applied but with 200 km/h, the application of GSM-R and introduction of ETCS Level 2 allows

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 67–74, 2010.

to implement the CBD rule with 250 km/h speed. (For truck trains the CDB is restricted to the same 120 km/h speed).

- For analysis the absolute RDB model is assumed: the zero time for previous train stop after sending the last message about position (the constant protect way for breaking).

2 The Changeable Block Distance

The Fixed Block Distance (FDB) is a control method of proper sequence of trains is now based on positioning with respect to whole length between neighbor stations (so called one-distance block system) or with respect to block distance (so called multi-distance block system with assumed number of sections). Such method allows to localize the train in the one section of the block system (1300-1500m) according to the accessible speed of the train.

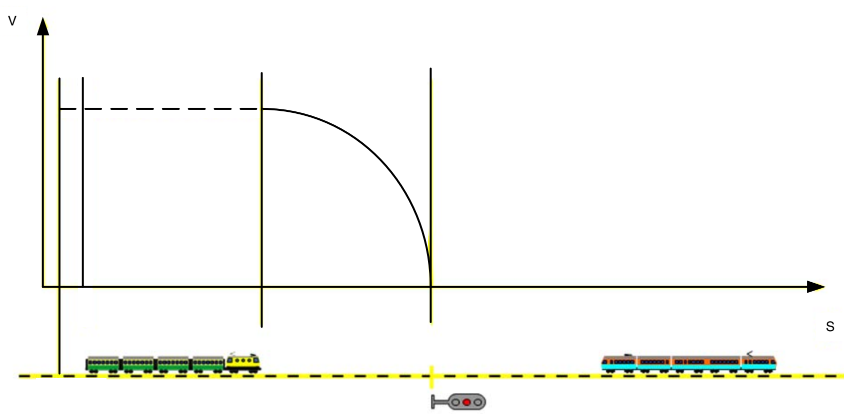


Fig.1. The rule of Fixed Block Distance (FDB)

This method assumes (in rough approach) division of the length between stations into given number of insulated sections related to required breaking way of the train. The occupation of the train the section after block signal transmits the message "occupied" (red light) to the this block signal and change information on a previous block signals, at the same informing the next train about number of not occupied sections and in consequence about permitted speed.

For ETCS level 1 (without GSM-R) the distance between neighbor balises may be similar to appropriate block sections because they must be placed near signals. It means that capacity of trains will be on the similar level (the increasing of speed from 160 km/h to 200km/h decreases the occupation time of rail section about 10%, but with assumption on mixed traffic with truck trains such value will be rather small).

The rule of the Changeable Block Distance (CDB) presented in the Fig.2 bases on the virtual block section with not fixed reference points (connected with distance between station) but flexible modified corresponding to given traffic situation. This method may be compared with “electronic visibility” when the actual speed depends on position and speed of a previous train. We can distinguish two variants of CDB:

- Fixed Changeable Block Distance (FCDB) when the second train receives the permission of drive to the place nearer than calculated on last report about position of a previous train. (It is connected with assumption about “zero distance” stop immediately after sending the report and total breaking way with protection way of second train.)
- Relative Changeable Block Distance (RCDB) when the second train receives the permission (movement authority) to the place nearer that the place whose may achieve end of a previous train after sending the last localization report. In such approach the distance between trains may be shorter than calculated breaking way of second train. In this method the very important is reaction time (delay in transmission of information about position of previous train end).

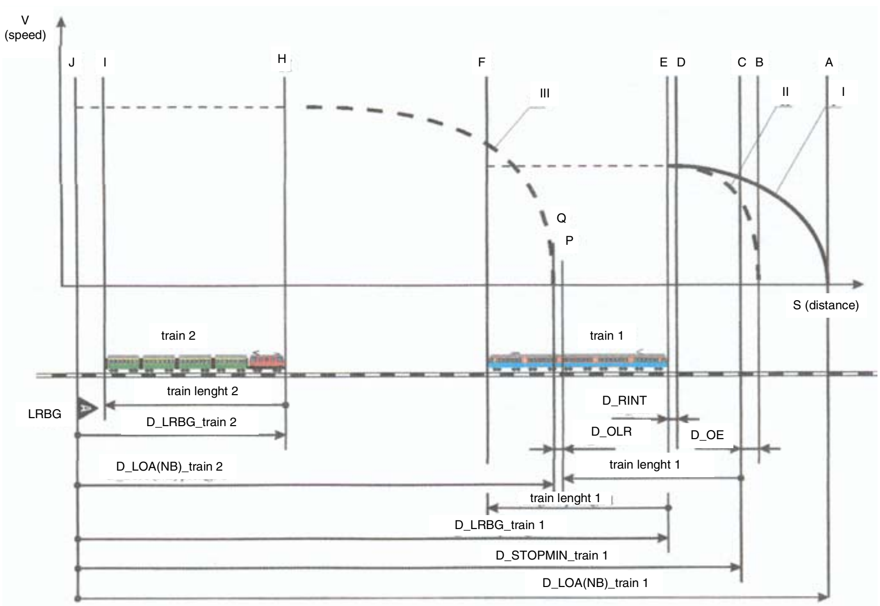


Fig. 2. The rule of Changeable Block Distance (CDB)

The distance control process is shown at the Fig. 2. According to this rule the ”train 2” receives Movement Authority (MA) about precise position with respect to calculated localisation of “train 1”. If the case of calculated distance between train 2 and 1 is shorter “train2” receive from Radio Block Centre new permission (MA) with limited speed and distance information. Finally when the driver outrun MA parameters train system automatically start – “ service brake – emergency brake”.

The introduced meanings and variables correspond to:

- MA – movement authority,
- LRBG – Latest Received Balise Group,
- I, II, III, - dynamic V profile, - breaking curves,
- A – end of movement authority for "*train 1*",
- B – calculated localization of stop point for "*train 1*",
- C – calculated safety point of stop for "*train 1*",
- D – theoretical localization of train 1 (situation when RBC indicated information – the last localization report of train 1 lost)
- E – latest localization of front "*train 1*" – according to LRBG information,
- F – latest localization of end of "*train 1*" – according to LRBG information,
- H – latest localization of front "*train 2*" – according to LRBG information,
- I – latest localization of end of "*train 2*" – according to LRBG information,
- J – localization of LRBG (balises group),
- Q,P – safety point for "*train 2*" (stop point).

In such application (with GSM-R) the minimal distance between Train_1 and Train_2 may be assumed as:

$$I = I_{h2} - I_{h1} + s \tag{1}$$

where:

I – minimal safe distance between trains equals to maximal breaking way (with additional safety margin)
I_{n1} – breaking way of previous Train_ (with fixed distance equals to zero)
I_{n2} – breaking way of next train
S – safety margin.

The cover of GSM-R for existing railway line requires the 99.9999% of availability for data transmission (in both full duplex channels). Such requirement assumes:

- The distance between base stations, (BTS) is 3-4 km (maximum 4 trains in 2 directions assigned to 1 BTS for CMK line).
- The radio control/monitoring centre (RBC) may service up to 40 trains.
- The time necessary for processing of information from balise and train odometer is 200 ms.
- The processing time of telegram in RBC is 200 ms.
- The time necessary for connection train a – RBC is 400 ms (3 correct telegrams each 400 ms required), the totals time is about 2 s.
- The time necessary for transmission of telegram train – balise is 200 ms.
- The report and transmission of information about position (each per several seconds) may be assumed as continuous, the deck computer calculates the position with respect to last reference/zero point. These calculations 200 ms plus connection/sending the telegram 2000ms gives the whole cycle of transmission about 3 s including additional delay and processing.

- The time and frequency of transmission the permission to the train depends only on the periods when the train traffic situation changes (according to actual permission of drive and required time of inspire.

3 The Object and Connected Characteristics

For the CMK as reference line we can assume the following characteristics of insulated sections: the shortest – 1200 m, the longest – 2400 m, the medium – 1500m. The detailed analysis of all block sections and connected signals suggests that in existing structure the required distances of breaking way for trains with 200 km/h may be rather difficult to achieve.

For partial analysis the typical distance between neighbor station (Psary – Góra Włodowska) is shown in the Fig. 3

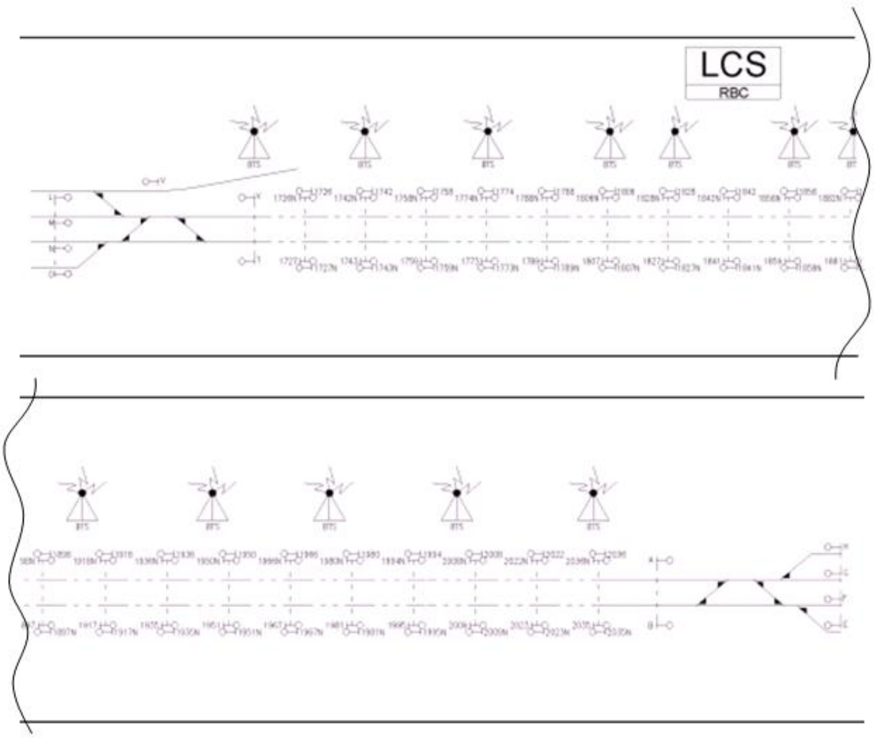


Fig. 3. The example of distance between two neighbor stations for CMK line

The detailed characteristics of traffic in the object from Fig. 3 is presented in the following table. This distance has length 34.715 km, identical for both directions of rails.

Table 1. The time table for object from Fig.3

	Block no.	km	[m]	Block no.	km	[m]	Block no.	km	[m]	Block no.	km	[m]
Station	**Track 1** (Grodzisk-Zawiercie)			**Track 1** (Zawiercie-Grodzisk)			**Track 2** Grodzisk-Zawiercie)			**Track 2** (Zawiercie-Grodzisk)		
Psary	B	168630	1319	F	169378	1899	A	168630	1319	G	169406	1719
Track 3 / Track 4:	**(6 Switch / *8 Switch*)**			*G*	*169406*	*1871*	**(7 Switch / *11 Switch*)**			*H*	*169547*	*1578*
Track 3 / Track 4:	*O*	*170256*	*1626*	**(6 Switch / *8 Switch*)**			*L*	*170220*	*1590*	**(7 Switch / *11 Switch*)**		
	N	170256	1626	T	171277	1426	M	170318	1688	V	171125	1581
	1727	172673	2417	1727N	172703	1560	1726N	172673	2355	1726	172706	1557
	1743	174233	1560	1743N	174263	1597	1742N	174233	1560	1742	174263	1597
	1759	175827	1594	1758N	175860	1474	1758N	175827	1594	1758	175860	1474
	1773	177304	1477	1773N	177334	1548	1774N	177304	1477	1774	177334	1548
	1789	178849	1545	1789N	178882	1790	1788N	178849	1545	1788	178882	1822
	1807	180672	1823	1807N	180672	2119	1806N	180672	1823	1806	180704	2087
	1827	182760	2088	1827N	182791	1420	1828N	182760	2088	1828	182791	1420
	1841	184180	1420	1841N	184211	1645	1842N	184180	1420	1842	184211	1645
	1859	185825	1645	1859N	185856	2278	1858N	185825	1645	1858	185856	2278
(21 block)	1881	188104	2279	1881N	188134	1651	1882N	188104	2279	1882	188134	1655
	1897	189754	1650	1897N	189785	2022	1898N	189754	1650	1898	189789	2018
	1917	191777	2023	1917N	191807	1801	1918N	191777	2023	1918	191807	1801
	1935	193578	1801	1935N	193608	1451	1936N	193578	1801	1936	193608	1451
	1951	195029	1451	1951N	195059	1595	1950N	195029	1451	1950	195059	1595
	1967	196632	1603	1967N	196654	1470	1966N	196632	1603	1966	196654	1470
	1981	198093	1461	1981N	198124	1368	1980N	198093	1461	1980	198124	1368
	1995	199462	1369	1995N	199492	1405	1994N	199462	1369	1994	199492	1405
	2009	200866	1404	2009N	200897	1401	2008N	200866	1404	2008	200897	1401
	2023	202266	1400	2023N	202298	1311	2022N	202166	1300	2022	202298	1311
	2035	203579	1313	2035N	203609	2161	2036N	203579	1413	2036	203609	2231
Góra Włodowska	B	204939	1360	F	205770	1464	A	204939	1360	G	205840	1394
The minimal block:		1313	m		1311	m		1127	M		1311	m
The maximal block:		2417	m		2278	m		2355	M		2278	m

4 The Capacity Evaluation

The capacity of railway line is exploitation parameter describing the maximal number of train pairs passing the railway line (or fragment) in assumed time (day or rush hour). It depends on following elements:

- Time-table speed
- Number of rails
- Length of rails and Block distances
- Composition of station rails
- Application of railway control devices
- Structure of traffic (type of trans)

From many different methods of capacity calculation the analytic method determines the element with minimal capacity (so called critical element). The time of occupation (or blocking with respect to UIC 406) of rails by two trains with sets of routes are similar in the traffic chart:

$$T = t_1 + t_b + t_2 + t_a \tag{2}$$

where:

t_1,t_2 – time of driving through rail (in both directions),

t_a, t_b – time calculating in A and B stations.

The minimal capacity of rail section in pairs of trains per day may be expressed as:

$$N = (1 - \varphi) * \frac{1440}{T} \tag{3}$$

where:

T – blocking time of section (traffic chart period)

φ – coefficient of traffic flexibility (always from period 0.2 – 0.3)

Assuming the parameters of FBD and existing control command systems (max speed 160 km/h) we can calculate medium time occupation block – 23 min and the capacity as 24 trains per day (maximal theoretical capacity as more 70 trains per day).

The maximal capacity ensuring by FBD method with ETCS L1 equipment for this object is connected with the following parameters: maximal speed 250 km/h and additional 5 HS trains per day. The calculated medium time of occupation block is 21 min and the capacity 29 trains per day (maximal theoretical capacity about 90 trains per day).

The application of CDB method (trains with 250 km/h and 4100 m distance - breaking way between trains 3500m plus train length 600m) allows to achieve the better parameters and in consequence the optimal value of occupation block (16 min) and capacity (29 trains similar to FDB and the maximal theoretical capacity about 100 trains per day). For traffic with HS trains only with 250 km/h theoretical maximal capacity may be 8 trains per hour or 200 trains per day.

5 Conclusions

It is obvious that virtual positioning of train according to GPS position may increase the capacity of trains. Assuming estimated value of capacity and assumptions on delays in transmitting message we can assure, that existing structure is sufficient for safe supervision of trains in CMK line, because the required number of trains assigned to BTS and RBC is satisfied.

References

1. Lewiński, A., Perzyński, T., Toruń, A.: The modeling of data radiotransmission for ERTMS application. In: Mikulski, J. (ed.) Advanced in Transport Systems Telematics. Faculty of Transport Silesian University of Technology, Katowice (2009)
2. Pawlik, M.: Metoda indykacji pojazdów w systemach sterowania ruchem kolejowym. Rozprawa Doktorska – Politechnika Warszawska Wydział Transportu, Warszawa (2001)
3. Toruń, A.: Wireless data transmission systems as a source of train localisation information for Signalling and Traffic Managment Systems. In: Mikulski, J. (ed.) Advances in transport systems telematics, pp. 327–334. WKŁ, Warszawa (2008)
4. Toruń, A., Lewiński, A.: Informacja przestrzenna w procesie sterowania ruchem kolejowym. materiały konferencji LOGITRANS 2010, Politechnika Radomska 2010, Logistyka 2/2010 (płyta CD) (2010)
5. UIC Code 406 "Capacity". UIC, Editions Techniques Ferroviaires, Paris (2004)
6. Wendler, E.: ETCS und Kapazität (ETCS and capacity). In: Proc. VDE Kongress 2006 Aachen, str 369-374, vol. 2, VDE-Verlag, Berlin (2006)
7. Wendler, E.: Weiterentwicklung der Sperrzeitentreppe für moderne Signalsysteme (Further development of blocking-time sequences related to modern signalling systems). Signal+Dracht 87, 1995/7-8, str 268-273
8. Gorczyca, P., Mikulski, J., Białoń, A.: Wireless local networks to record the railway traffic control equipment data. In: Advances in Electrical and Electronic Engineering, Żilina, pp. 128–131 (2006)
9. Młyńczak, J.: Lokalne centra sterowania w warunkach polskich. Infrastruktura Transportu 5, 55–58 (2009)

Implementation of Network Redundancy in Environment of Road Tunnel Control

Anna Cerovská[1] and Juraj Spalek[2]

[1] BETAMONT, s.r.o., J. Jesenského 1054/44, 960 03 Zvolen, Slovakia
cerovska.anna@betamont.sk
[2] Faculty of Electrical Engineering, University of Žilina,
Univerzitná 1, 010 26 Žilina, Slovakia
juraj.spalek@fel.uniza.sk

Abstract. Availability of individual segments of a road tunnel control system is directly dependent on network infrastructure. A safe tunnel requires a 100 % availability of control. One of the alternatives how to approach the fulfillment of this requirement is the implementation of hardware and link redundancy to the network topology of the road tunnel by means of mechanisms that enable to improve the failure tolerance in Ethernet networks.

Keywords: Tunnel backbone, redundancy, network, redundancy protocols.

1 Introduction

For many current intelligent traffic systems a 24-hour service is required at every of 365 days of the year and to provide this is not so "easy" [4]. Since each system or transmission medium has its life cycle and failure rate, it is necessary to explore this factor. One of the alternatives can be using of redundancy. Its function in the system is to react in desired way to the failure. For that reason it can serve for recovery after the failure. The reserve parts of the system are those, the using of which would be obsolete, if other parts of the system work correctly. According to the way of using reserve in a specific time, we can recognize 2 types of reserves [3]:

- Active reserve – all facilities that perform a certain function are determined for current operation (in technical literature designated also as a "static reserve");
- Standby redundancy – a part of used facilities performing the required function is determined for the working, whereas other parts of used facilities do not perform any work as far as it is not necessary (in technical literature designated also as a "dynamic reserve").

However, systems containing active reserve and standby redundancy often occur in professional praxis, the so called hybrid redundant systems.

Another way to differentiate redundancy is according to elements which require backup:

- **Power redundancy** - industrial products should have at least two power inputs to accept power from a primary and a backup source, guaranteeing uninterrupted operation.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 75–84, 2010.

- **Media redundancy** - A basic requirement for industrial networks is media redundancy, which involves forming backup paths for network access.
- **Node redundancy** - In many industrial networks, certain devices must always be available and communication must not be interrupted at any time, otherwise great losses are incurred. For this reason, critical devices can be backed up by setting up dual network nodes. Both network nodes should be connected to a dual-homing controller, which is able to select the most suitable homing path. To continue normal network communication even when a network disaster occurs, the dual-homing control must establish connections with certain critical end devices.
- **Network and system redundancy** - Some industrial networks may rely on two physical networks, even two complete systems, as a redundant solution. Once media and node redundancy have been implemented, advanced management of redundant systems must be taken into consideration, including the management of two completely independent networks with two communications ports on each connected devices. Network and system redundancy are more complete solutions, but involve greater cost and complexity [2].

2 Control of Safety Critical Processes

Control system of road tunnels can be designated as control of safety critical processes (i.e. safety relevant systems) which fulfill the control functions with defined safety level. For safety critical process such technological process is considered, at which the failure of prescribed functions can bring up the occurrence of non-intended event, or sequence of events can incur that cause endangering of human lives, significant damage of environment, large material losses, or may cause the loss of provided service [1].

Safe road tunnel requires 100% control availability. Availability of individual segments of the road tunnel control system depends directly on network infrastructure. The uppermost task at its designing is the proposal of appropriate network topology. To achieve the best possible availability of network segments by simple doubling of all active network components and interconnections would be very costly. Therefore it is appropriate to analyze carefully each control level and to evaluate the necessary stage of network redundancy.

3 Network Redundancy Mechanisms

In recent years, redundant Ethernet technology has been rapidly and popularly adopted in the industrial automation field due to its enhanced reliability. There are many existing mechanisms that can enhance fault-tolerance in an Ethernet network. The most common of these are the Mesh networking, STP, RSTP and proprietary ring redundancy [2].

Mesh networking is a network where all the nodes are connected to each other in a complete network. Data travelling on the Mesh network is automatically configured to reach the destination by taking the shortest route. Mesh networking is reliable and

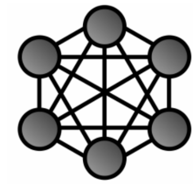


Fig. 1. Full meshed network

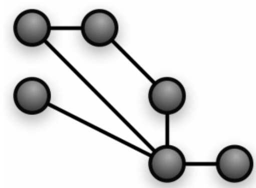


Fig. 2. Partially meshed network

self-repairing. If one node fails, the network finds an alternate route to transfer the data, but there are increased costs from the cables required to connect all the nodes to each other. See Fig. 1 and Fig. 2.

Spanning Tree Protocol (STP) as defined in the IEEE 802.1D standard is designed to eliminate loops in a network by cutting the network into a loop free tree shape. It helps a network to achieve link redundancy and path optimization. Its functionality rests upon the fact that between two nodes only one route can be active. Several active routes between nodes can cause loops in network and those data duplication. In spite of the fact that STP resolves loops in network at retaining redundancy it also has its bad features. One of them is a long time for network renewal, usually up to 15 seconds as far as the spanning tree is stabilized and this time is too long for industrial applications.

Rapid Spanning Tree Protocol (RSTP) – to overcome the slow convergence of STP, the IEEE released the IEEE 802.1w standard to make improvements based on STP. This shortened the recovery speed to 1 second. However, for some real-time and mission critical industrial applications, such as road tunnel control system, recovery time must be under 100 milliseconds to ensure the reliability of the network. STP and RSTP are open standards that many Ethernet switch manufacturers have implemented in their managed switch products. The faster self-healing time of RSTP is very helpful in an enterprise network where a few seconds of network delay is acceptable. However in an industrial control network, one second of missed communications can cause serious problems.

Look at visual explanation of basic functionality of Spanning Tree Protocol in Fig. 3.

Ring redundancy is common among today's industrial Ethernet networks. It is a more cost-effective solution than a mesh network, and overcomes the recovery time problem of STP and RSTP. There are many different ring redundancy technologies featuring a guaranteed recovery time of a few milliseconds offered by industrial Ethernet solution providers. Ring redundancy ensures non-stop operation of networks

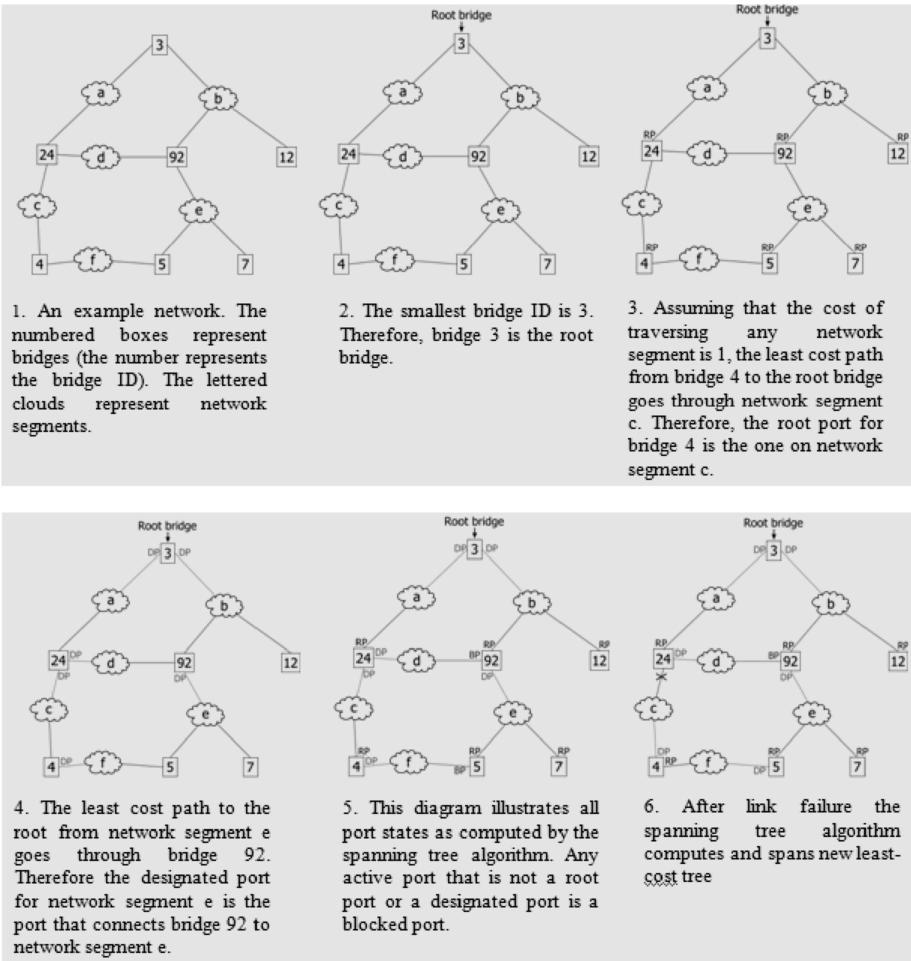


Fig. 3. Spanning tree protocol principle of operation

with an extremely fast recovery time. If any segment of the network is disconnected, the network system will recover in a few milliseconds by activating the backup path in a ring. It is, however, necessary to consider the number of connected network nodes in one ring. With higher number of nodes there is also increase of probability that some nodes may fail at the same time which causes unavailability of those nodes that are connected among the failed ones. It is applicable to create smaller network rings. There are several possibilities how you can connect network rings each other (see Fig. 6). Nevertheless, the ring topology has some significant advantages – mainly an acceptable ratio redundancy level / price which moves it into the position of preferred network topology in the area of intelligent control systems [2].

A completely redundant system consists of redundant switches, redundant communication ports, and redundant device pairs. All Ethernet devices and workstations are connected to both independent ring network architectures (see Fig. 4 and Fig. 5).

Depending on the circumstance, there are two possibilities that fit this redundancy application. One of the possibilities uses devices that have two ports, with one of the ports utilized for the primary path, and the other port serving as the secondary path. The other possibility uses devices that have only one port. In this case, the devices must be upgraded to two Ethernet ports, in order to form the primary and secondary paths (Fig. 6).

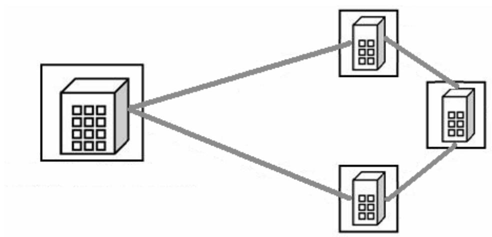


Fig. 4. Ring type network

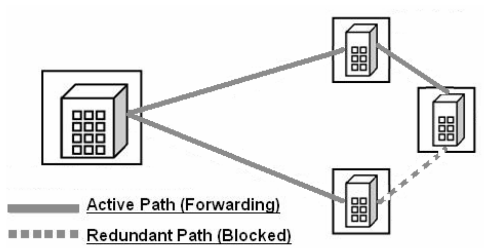


Fig. 5. Ring type network with one segment blocked

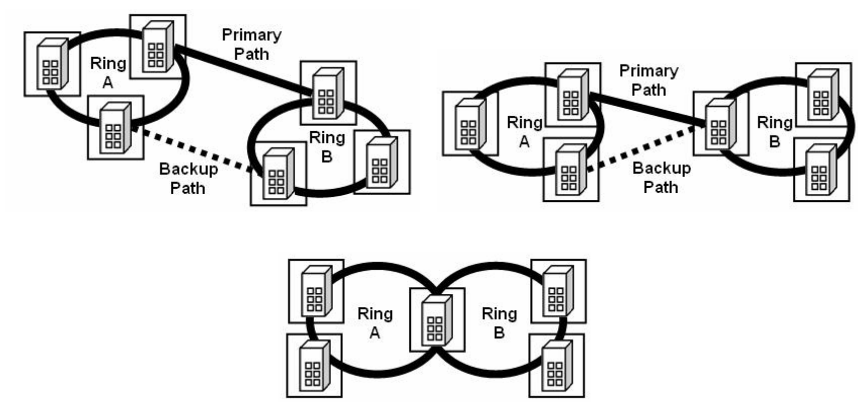


Fig. 6. Ring – to – ring applications

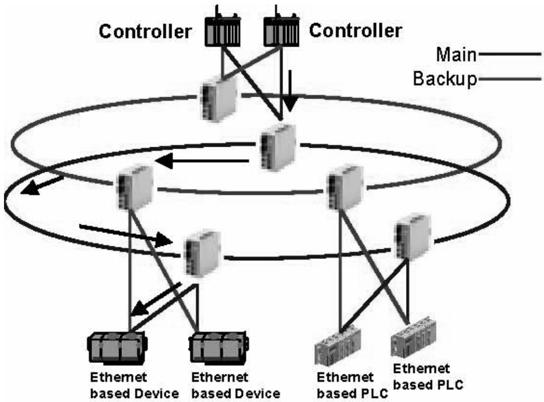


Fig. 7. General flow control in two independent ring network architectures

Complete system redundancy can form an extremely reliable network that minimizes data loss and has fast recovery time (Fig. 7). There must be a dual homing controller that is able to distinguish which Ethernet device is active the primary path or secondary path. The diagnostics can ensure that active devices are fully functional and ready to take over at any time (see Fig. 8 and Fig. 9) [2].

4 Transmission Mediums

The used transmission medium also plays important role in process of road tunnel control. Nowadays the using of fiber optics is preferred since the fiber optics has in comparison with traditional metallic leading couple of benefits, first of all a large bandwidth that gives a possibility to transfer a larger data amount. Signal passing via fiber optics is also significantly less inclined to the disturbances; fibers are much thinner and lighter than classical wires. Cupper line has its limits both in bandwidth and length of line which is at the present time not sufficient.

Optic fibres are basically divided into single mode and multimode ones. The single mode fibre is characterized by its thin core. It is less sensitive to attenuation than multimode fibre, has higher performance and spreads only to one direction with high concentration of photons that allows transmission of information on longer distances (max. distance 3 km) in high speeds. The multi mode fibre is thicker than the single mode one which causes that the light beam can spread in more paths that provides a larger bandwidth, however, at the lower transmission speeds. As a source of light beams there are used the LED-s. It is suitable for realization of interconnections on shorter distances (within 2 km).

The tunnel construction is characterized with its short distances between the connecting points and therefore it uses at the realization of backbone network often exactly the multi mode fibre optic cable which is more acceptable also from economical point of view. It is the price for the cable itself and necessary signal converters (converting the electric signal to the light one and vice versa) that plays the decisive role. The single mode fibre optic cable is often primary used on open motorway sections where the density of the technological components connected to fibre optic line is lower.

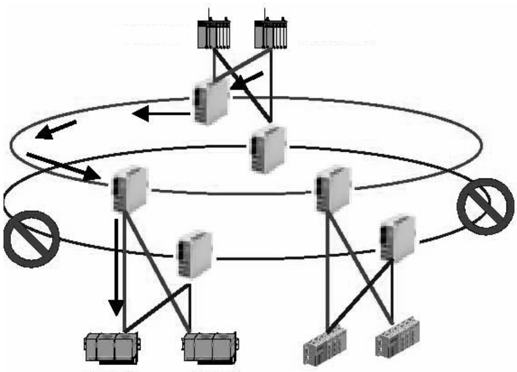


Fig. 8. Network failed

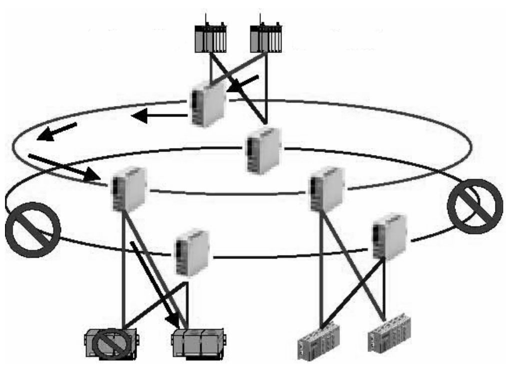


Fig. 9. Network and device failed

5 Constructional Readiness for Redundancy Implementation

Constructional architecture of tunnel constructions plays also important role at the proposal of redundancy for road tunnel backbone network. It affects mainly the problems of placing the redundant elements of network where a physical re-distribution of safety critical elements on various places is preferred. For instance, it is appropriate to lead the alternative communication routes through a separated cable line in order that the availability of communication is secured also in case of damaging the primary cable route. Similarly, it is appropriate to arrange two local operational centres for a road tunnel (control rooms) with placing of central control system of the particular road tunnel in order that the superior level, for example a regional operational centre (which is in charge of control and supervision of all tunnels in region), or a central operational centre (administration and supervision of all road tunnels in Slovakia) has in case of a failure of one of them an alternative access to the control. It is not necessary that both local dispatching centres are equipped with surveillance for operation (CCTV surveillance, operator working station etc.), since the control command comes from the superior level. It is sufficient just to place the redundant technological elements in ideal way, for instance on both sides of the tunnel tube.

6 Proposal of Redundant Backbone Network for a Selected Tunnel as an Example

Cable lines in this model tunnel (dashed lines lines in Fig. 10) are proposed by both alongside of both tunnel tubes to secure the physical separation of alternative communication paths. Technological and at the same time network nodes are in this case placed in special electrical distribution rooms (or cabinets) in cross interconnections of tunnel tubes (dark boxes in Fig. 10) that serve as escape ways both for persons and vehicles of emergency crews.

The redundancy on this level is provided in the form of doubling of nodes whereas one node serves primary for mediation of control system communication with control segments for a certain section of the first tunnel tube (e.g. the southern tunnel tube) and the second one serves primary for mediation of control system communication with control segments for a certain section of the second tunnel tube (e.g. the northern tunnel tube). They are, however, interconnected, in order to provide a substitution of these nodes if one of them fails. On both sides of the one, or the second tunnel tube are located redundant control rooms of the tunnel where also a connection point to the main fibre optic cable route of the whole motorway section is situated. By doubling the tunnel control room we secure the availability of central control system also if one of these control rooms fails (failure, or fire).

Active network elements[1] are connected into a ring whereas primary communication path is led through internal cable line of one of the tunnel tubes and the network ring is closed by outer cable line of the second tunnel tube. That means it is provided that except of the fact that the alternative communication path is led through another cable line, the cable line is in other tunnel tube as well. In such a way two primary network rings connected in level of control rooms into two redundant active network elements (each ring is terminated in another active network element) are created. It is still possible to increase the network availability by doubling these proposed network rings in that on the level of control room they will be connected into another active network element as it is at primary network rings (cross connection). See Fig. 11.

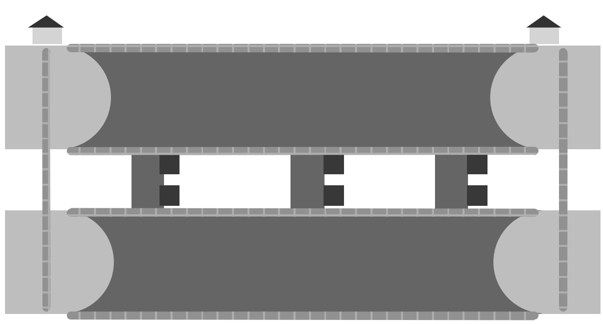


Fig. 10. Cable lines

[1] In this case network switches, eventually routers are assumed

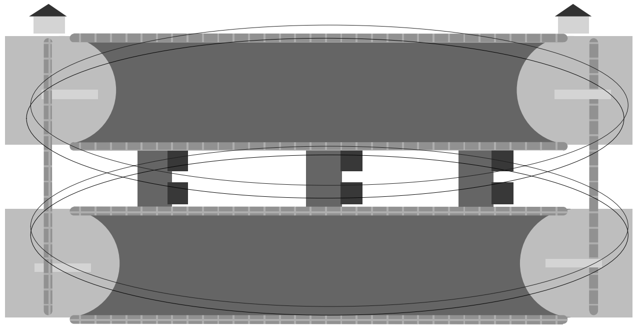


Fig. 11. Network rings

But this is only one variant of backbone network interconnection. As it was mentioned in chapter about ring redundancy, it is more reasonably to create smaller network rings, mainly in case of long road tunnel. There is possibility to divide tunnel tube into more sections connected each other.

The goal of this complex network topology is to achieve the control availability also in the most critical cases, such as fire in control room and the following damage of the placed technology, or fire in one of the tunnel tubes, eventually to create a possibility to accomplish service without the necessity of disconnection the other parts of the control system.

7 Mastering the Failure Consequences

Not even the careful application of proceedings is able to exclude completely the occurrence of failures during system operation. Thus, the system has to provide such features that enable to find out a failure and to guarantee the transition of the system into a safe default state. That means for instance that if one part of the control system (control segments) losses connectivity with central control system, those control segments will locally control the traffic process according to the defined default scenarios. If a failure of the control segment itself occurs, an initial safe level has to be predefined into which the individual controlled technological elements would get in case of detected fault. Re-leaving of this state can occur only after failure removal. Repeated putting the system into operation cannot occur spontaneously, but must be under control.

8 Conclusions

The goal of this paper is to demonstrate to a reader the problem of availability improvement in the safety critical technical systems situated in road tunnels. On the principal level basic techniques and mechanisms for success redundancy implementation in road tunnel backbone network were explained.

Acknowledgements. This work has been supported by the Grant Agency VEGA of the Slovak Republic, grant No. 1/0023/08 "Theoretical apparatus for risk analysis and risk evaluation of transport telematic systems". The paper was prepared under the support of the ERDF - European Regional Development Fund, project No. ITMS-26220120028, "Centre of excellence for systems and services of intelligent transport".

References

1. Zahradník, J., Rástočný, K., Kunhart, M.: Safety of Railway Interlocking Systems, pp. 27–30. University in v Žilina /EDIS – publishing house ŽU, Slovak republic (2004) ISBN 80-8070-296-9
2. CCNA Exploration 4.1, Cisco Networking Academy Course Materials
3. Přibyl, P., Janota, A., Spalek, J.: Analysis and Risk Control in Transport. Technical literature BEN, Praha, pp. 419–432 (2008) ISBN 978-80-7300-214-5
4. Urgela, S.: Applications of the Railway Information Systems (in Slovak). In: Road Conference, pp. 218–223 (2003)

Concept of "One Window" Data Exchange System Fulfilling the Recommendation for e-Navigation System

Damian Filipkowski and Ryszard Wawruch

Gdynia Maritime University, Faculty of Navigation,
Morska Str. 81-87, 81-225 Gdynia, Poland
{dfilipkowski,wawruch}@am.gdynia.pl

Abstract. The implementation in maritime radio-communication of so called "One window concept" for exchange of information between a ship and a port and coastal state authorities requires designation of one contact point on shore for these purposes, e.g. harbour master or ships' monitoring or traffic control centre. In Poland, as contact points regional and local centres of the Polish National Maritime Safety System will be designated. Paper describes the proposal for system of data exchange between a ship and a shore contact point, containing definition, functions and architecture of proposed system, possible directions of information flow and levels of access, fulfilling requirements of this concept and recommendation for e-navigation system developed on the basis of the International Maritime Organization (IMO) and International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA) working papers.

Keywords: Maritime radio communication, e-navigation, one window concept.

1 Introduction – The Definition of e-Navigation

The International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA) has formulated idea of e-navigation several years ago. Current trends in maritime activities to increase navigational safety, security, efficiency, responsibility and environmental protection should include expected development of technology and result in significant changes in shipping and navigation. Therefore, it is necessary to be reasonable and universal integration of navigation systems (devices) and services is intended to serve minimize navigational errors, accidents and incidents, improving safety and reducing the costs for shipping and maritime administrations. For work of IALA, quickly joined by such organizations as the International Maritime Organization (IMO) and the International Hydrographic Organization (IHO), which in accordance with its main tasks, operate to systematically maintain the required level of safety of human activities at sea [1].

During its works IALA formulated the definition of e-navigation. Later, the IMO has modified the definition by adding a few words, and so was the official definition of e-navigation system as follows:

"E-Navigation is automatic creation, collection, management and exchange of maritime information at an adequate level of secrecy of the information, and minimal user

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 85–95, 2010.

intervention, using electronic means, on board and ashore, through the one contact point, in order to increase the level of safety of navigation, improved navigation and the wider marine environment" [2], [3].

This definition contains all the elements, which characterize the system of e-navigation, and extracts its basic functions. To make this definition complete there could be added that information cannot be freely available and the user role should be emphasized. It could also be stressed that the primary function is to increase the safety of navigation, rather than its provision, because a level of safety is ensured and place this function first, to emphasize that this is a basic function. There also should be added that the system should be as far automatic as possible. According to the proposed amendments to the definition of e-navigation would be as follows:

"E-navigation is the harmonised creation, collection, integration, exchange and presentation of maritime information onboard and ashore by electronic means to enhance berth to berth navigation and related services, for safety and security at sea and protection of the marine environment."

Described researches were conducted in the scope of research work financed by the Polish Ministry of Science and Higher Education as developmental project from the means for science in 2008-2010 years.

2 The Architecture of e-Navigation

In this chapter there is presented the proposed architecture of an e-navigation system. Fig. 1 and Fig. 2 illustrate possible flow of information and show that data may be transmitted simultaneously in different directions between ships and between ship and shore. As shown in these diagrams, information exchanged by the system will have to be additionally protected against disclosure or theft. Lines in Fig. 1 and Fig. 2 mark all the possible directions of information flow between two ships and between ship and VTS station. Solid lines mean that the user does not request the information. Long-dashed lines regard data on which the user has expressed demand and obtain access to its, and short-dashed lines mark it with the information which the user expressed the need, however, did not have sufficient rights to receive it.

Shortcuts used on diagram:

BAMS - Broad Area Maritime Surveillance;
ISPS - International Ship and Port Facility Security Code;
MAS - Maritime Assistance Services;
PSC - Port State Control;
SAR - Search and Rescue;
FAL - Facilitation of International Maritime Traffic Convention;
VTS - Vessel Traffic Service;

VTS (Vessel Traffic Service) needs the following information about the vessel:

- Identification;
- Position;
- Speed;

- Course;
- Draught;
- Passage plan;
- Navigational status;
- Expiry dates of ship's certificates;
- FAL documents;
- Information about the passengers and crew;
- Information about cargo on board and port of discharge;
- The last 10 ports visited (in accordance with ISPS- International Ship and Port Facility Security Code);
- Information about last PSC control;
- Sanitary information;
- BAMS information;
- Information about SAS (Security Alert System); and
- Information about Security Plan.

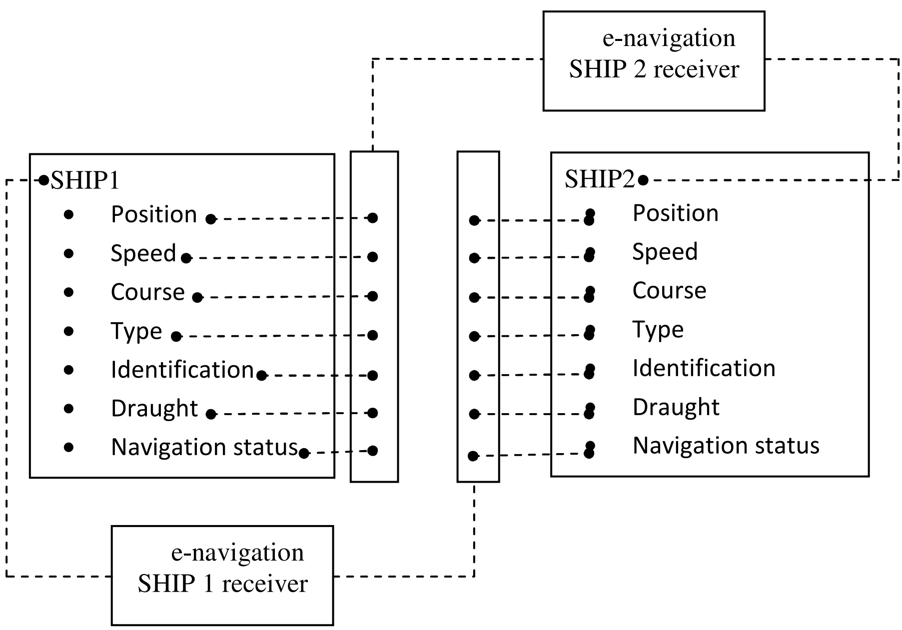


Fig. 1. Possible information flow in e-navigation between two ships

However, there is the possibility that VTS has not been allowed to obtain data included in the sip's Security Plan (short-dashed line on the Fig. 2).

For example, at the same time the ship must update its travel plan, and requires the following information:

- Position and time of pilot embarkation;
- Weather forecast;

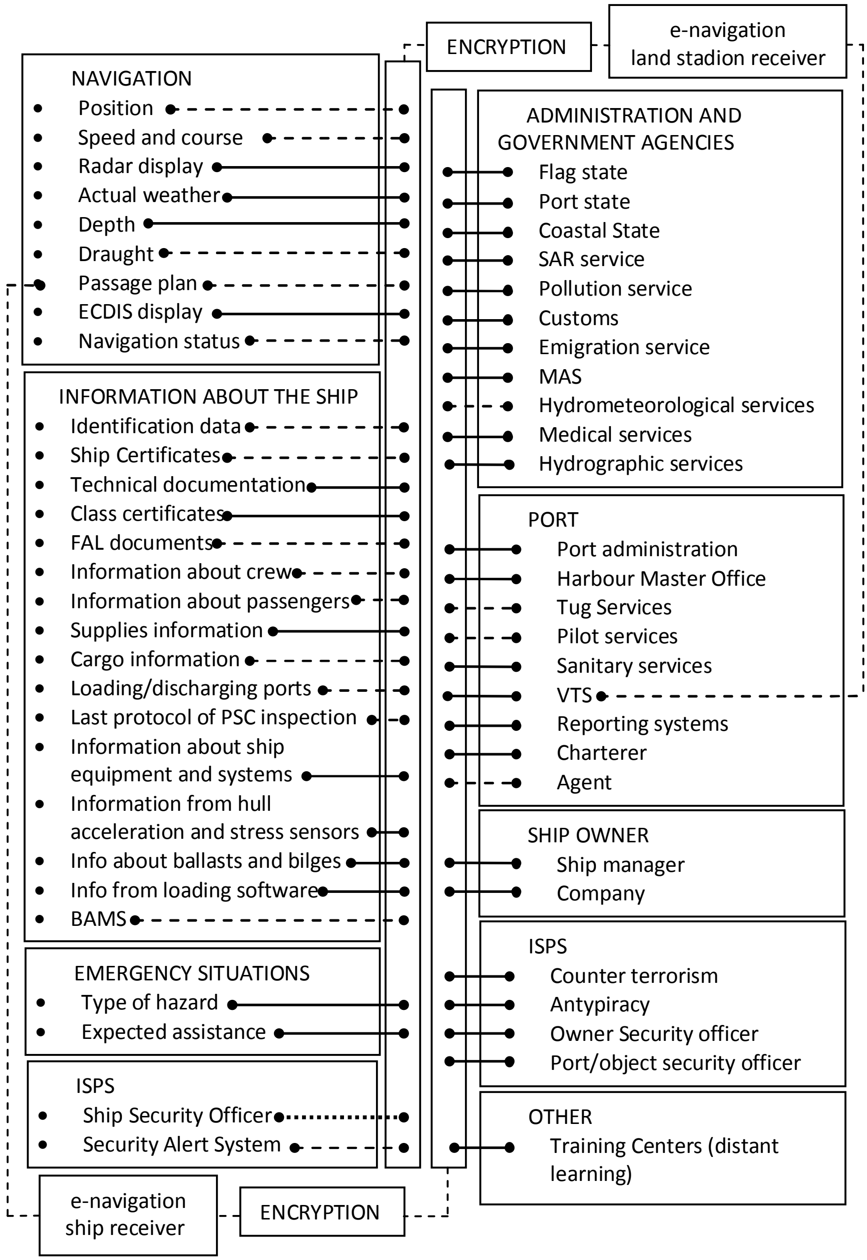


Fig. 2. Possible information flow in e-navigation between ship and VTS station on flow in e-navigation between ship and VTS station

- Tidal information; and
- Documents required by cargo receiver at the port of destination.

Ultimately, the ship receives the necessary information from the cargo receiver, towing services and hydrometeorological services and VTS only information to which it had access.

The system will storage the information (e.g. in computer database or VDR on vessel and in one contact point ashore) from sources like [4]:

- D/GPS (Differential/ Global Positioning System);
- AIS (Automatic identification System);
- Radar with tracking facilities (Automating Radar Plotting Aid - ARPA);
- ECDIS (Electronic Charts Display Information System);
- Navigation sensors; and
- Other.

Presenting the architecture of the system, there should be shown flow of data and profits that each user may receive. Below are some examples of such flow between vessel and shore.

2.1 Information Flow between Ship and Hydrometeorological Station

Fig. 3 shows possible data exchange between ship and hydrometeorological station.
Ships profits from these data exchange are as follow:

- More credible passage plan to allow to reduce travel time and fuel consumption actual weather messages; and
- Knowledge of actual:
  - Hydrometeorological conditions;
  - Information about tides and currents; and
  - Weather forecasts.

All this information received in accordance with the assumptions should be presented in an accessible way (e.g. chart, maps, text etc.).

Shore station receives as a profit weather data from the vessels, which may be used to create more accurate forecasts.

There should be emphasized one aspect of e-navigation system. Data stored on board, sent to the station and properly interpreted ashore, increases accuracy of information received from this station on ship back. This means that e-navigation system in this approach has the ability to upgrade information that vessel will receive.

2.2 Information Flow between Ship and VTS

Data possible to exchange between ship and VTS centre is presented on Fig. 4.
As ship’s profit from this data exchange may be mentioned:

- An additional source of data about traffic when approaching the harbour;
- Warnings and advice to the manoeuvres;
- Confirmation of time and position of pilot embarkation; and

- Sending all information without interference by the navigator (navigators can focus on running the ship).

Stations VTS profit are as follow:

- More effective traffic control;
- Possible increase of capacity of traffic in harbour; and
- Information about the ship with a large degree of assurance (no human error).

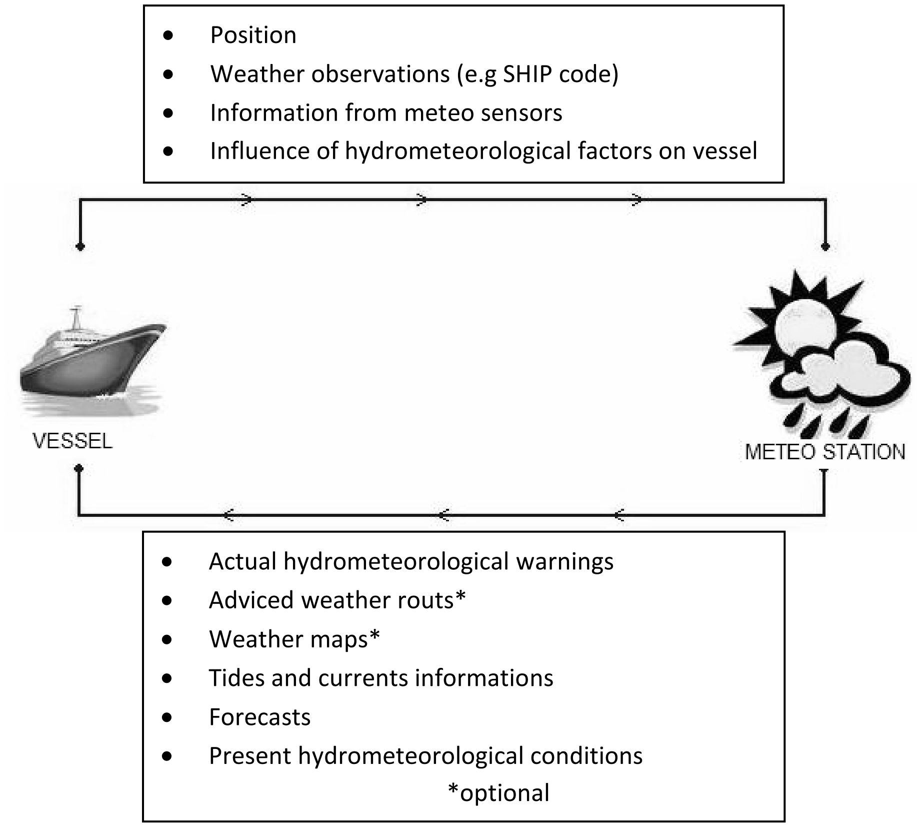


Fig. 3. Information flow between ship and hydrometeorological station

3 Tasks of e-Navigation System

Described system shall meet following goals [3], [5]:

- Increasing the safety and security of the ship;
- Streamlining and automation of ship-to-ship, ship-shore and shore-ship data exchange;
- More effective vessel traffic observation;

- More effective control of ships in terms of regulations;
- Increasing the efficiency of maritime transport;
- Improvement of SAR operations (access to the information without participation of vessel in distress);
- Increasing the accuracy of the information (by rejection of inaccurate and false data and the ability to compare information from a database on land);
- Reduction of shipping costs (e.g. fewer devices, lower insurance fees);
- Increasing the security of data transmitted (through secrecy and encryption of certain data);
- Minimalization of the likelihood of human error (automation);
- Possibility of division of responsibility (the pilot ashore will have access to the same information as the navigator on board);
- Maximalization of transparency of the presented information (by filtering unnecessary, not valid or false information and information with low degree of accuracy);
- The use of technological progress, taking into account the user requirements; and
- Pollution prevention.

4 Features of e-Navigation System

To meet the above tasks the system should have the following characteristics [3], [5]:

- Reliability on level of 99%-100% of work time;
- Integrity;
- Sufficient degree of redundancy to achieve established level of safety and security of the vessel;
- Ability to evolve with technological progress and changing user needs;
- Range sufficient to needs (global);
- Radio-communications subsystem and transmission of data with appropriate characteristics such as speed, reliability and range;
- High degree of automation; and
- Appropriate data encoding algorithm to obtain a high level of system security.

5 Ensuring Radio-Communication in e-Navigation

One of the biggest challenges for designers of e-navigation system will be creation of a radio-communication subsystem in such a way as to meet the goals defined for e-navigation. It is important that the radio-communication subsystem used in e-navigation is not only an integral part of that system but also retains a level of autonomy.

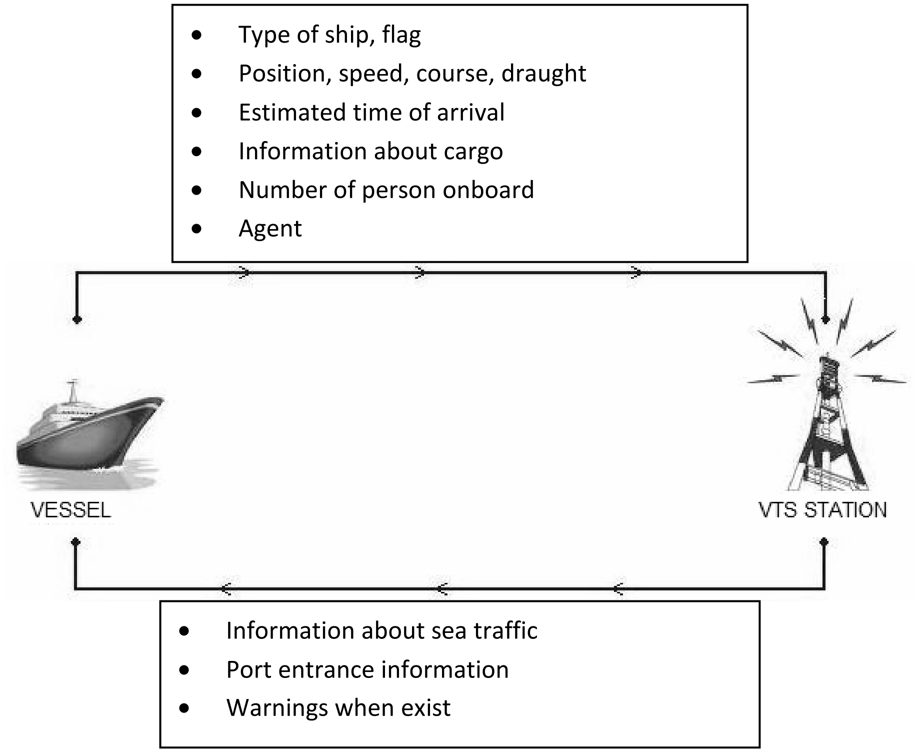


Fig. 4. Information flow between ship and VTS station

5.1 Functions of Radio-Communication Subsystem in e-Navigation

It is believed that present Global Maritime Distress and Safety System (GMDSS) is an archaic system and it has to be modified or exchanged. Designers of e-navigation radio-communication subsystem must consider whether it is worthwhile to modify existing solutions in terms of radio-communication, whether to create from scratch a completely new subsystem characterized by greater clarity, ease of use and the possibility of evolution than is currently used for GMDSS. The main function of GMDSS is to alert the distress and to liaise and should remain so. The only thing that should be changed is a way to alert, which should have simplified procedures. All information should be sent automatically. Sending alarm using two independent systems by pressing a single button is available now. Storage of valuable information in VDR or in ship's computer database and sending along with asking for help it is also possible with a high level of today's technology but the equipment of GMDSS doesn't give such a possibility. Creators should not just replace the existing GMDSS, but to go a step further and create a system not only functional (e.g. GMDSS works only between latitudes 70'N and 70'S, and new system should be global) but also friendly to user. As for other features it is dependent on what will be the final shape of the e-navigation [6], [7].

5.2 Functions of the Radio-Communication Subsystem

A radio-communications subsystem in e-navigation must provide access to the following services [6], [7]:

- Transmission of electronic mail (e-mail);
- Automatic updating of the navigational charts and publications in electronic form;
- Receiving of weather information;
- Access to web pages and intranet (e.g. master using company extranet);
- Secured communication;
- Transfer of large files over the Internet (e.g. pictures, video files etc.);
- Routine communications on the Internet and telephone;
- Transmission of SMS in relation ship to ship and ship to shore to ship (SMSs sent from mobile phones to vessel radio station);
- Videoconferences*; and
- Storing and sending video images*.

(*optional)

The easiest way to obtain these parameters by radio-communication subsystem to meet the goals defined earlier for e-navigation is to provide a range of services required by the official users (mentioned on diagrams) like, or similar to, Fleet Broadband. In addition to basic functions, Fleet Broadband provides a standards-based connectivity IP (Internet Protocol) and ISDN (Integrated Services Digital Network) and due to that it appears to be the most appropriate system for use with e-navigation. Additionally it provides the minimal requirements regarding security of data in radio-communication subsystem and offers a number of additional options for the user.

6 "One Window" Concept User Interface

Quick and easy access to information, integrity, reliability, high level of automation, possibility of warning in case of danger, ability to filter data and access to information are just some of the user interface features of an Integrated Bridge System. Presented software is an authors' suggestion of a "one window" concept user interface. The program was developed using object programming and running the Windows operating system, which allows users familiar with the Windows environment for easy handling of this program. Software used to create a program was Delphi 7 Personal freeware installed on computers in the computer laboratory of Gdynia Maritime University. Delphi 7 and Windows are not software, which can assure level of reliability mentioned before, but presented program is only an example and it was created to show principles of work of user interface.

User of e-Navigation will operate primarily on data that contains valuable information for him. To keep true value of information, "one window" concept user interface in e-navigation must have the following characteristics:

- Reliability;
- Integrity;
- Sufficient degree of redundancy to achieve the established level of safety;
- Possibility to evolve with technological advances and changing user needs; and
- Range of services sufficient for user needs.

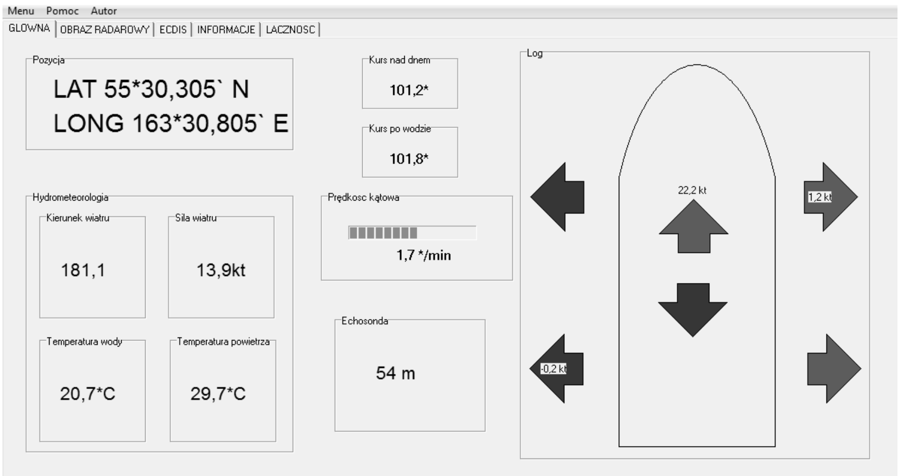


Fig. 5. Proposal of user interface

To let the e-navigation retained these qualities the following characteristics of the information must be defined in detail and fully specified:

- Availability;
- Timeliness;
- Fairness;
- Completeness;
- Comparability;
- Clarity;
- Processability;
- Cost efficiency (may be expensive or cheap);
- Addressability;
- Utility (the level of impact of the respective information on the effectiveness of decisions taken by the user;
- Prioritisation; and
- Value (the value of information is the difference between the utility of a decision by the decision maker under the influence and possession of the utility of a decision taken without holding that information).

The proposed software tries to meet the user needs while maintaining the above-mentioned features of information. Some of the data presented in the application is calculated using a mathematical function, and although some are generated by random algorithms entire application allowing to present a realistic simulation of the ”one window” user interface of an integrated bridge system.

The software works in Windows environment (it’s only an example and Windows has been chosen because it is the easiest operation system to work with). User has access to five tabs, each representing a different range of data presented to the user, and all together form an integrated whole:

- HOME tab - presents data that the user of an integrated navigation bridge can get and is presented on the main panel of IBS;
- RADAR DISPLAY tab - simulates the radar equipment installed on board together with the information it gives;
- ECDIS tab - presents data available for the user through the electronic chart navigation system IENC;
- INFORMATION tab - allows to present implementation of so called “One window concept” for exchange of data between ship and port and coastal state authorities and requires designation of one contact point on shore for these purposes; and
- COMMUNICATION tab – allows the user to simulate the process of establishing and proceeding radio-communication.

References

1. Nitner, H.: The role of hydrographic service in the development of e-navigation concept. Hydrographic review (6), 23–32 (2010)
2. Nordic Navigation Conference The IALA vision of e-navigation. Proceedings, Oslo (2007)
3. Sub-Committee on Safety of Navigation. Sessions 53–55. Working Papers. IMO, London (2007-2009)
4. Resolution A.861(20) Performance standards for shipborne voyage data recorders. IMO, London (2006)
5. Weintrit, A., Wawruch, R., Specht, C., Gucma, L., Pietrzykowski, Z.: An approach to e-Navigation. Coordinates III(6), 15–22 (2007)
6. Czajkowski, J.: System GMDSS. Skryba Sp.z.o.o., Gdańsk (2002)
7. Korcz, K.: GMDSS as a Data Communication Network for E-Navigation. Akademia Morska w Gdyni, Gdynia (2007)

Analytical Model of a Rail Applied to Induction Heating of Railway Turnouts

Elżbieta Szychta, Leszek Szychta, and Kamil Kiraga

Kazimierz Pułaski Technical University of Radom,
Faculty of Transport and Electrical Engineering,
26-600 Radom, Malczewskiego 29, Poland
{e.szychta,l.szychta,k.kiraga}@pr.radom.pl

Abstract. An idea of induction heating that may be useful in development of an effective method of heating railway turnouts in Poland is presented. Results of testing of a turnout induction heating method undertaken in 1978 – 1979 by the Polish National Railways operator (PKP) are discussed and provide the starting point for renewed research. A numerical rail model is introduced that is necessary to determine the distribution of magnetic induction and magnetic intensity across the rail itself and in its environment.

Keywords: Induction heating of turnouts, three-dimensional rail model, magnetic induction, magnetic intensity.

1 Introduction

Railway turnouts are key route elements, exposed as they are to such adverse weather conditions as snow, wind-swept snow masses, low temperature or sleet. It is therefore important to keep the turnouts operational despite these adverse weather conditions. Snow falling between the point and the rail (rheostat) and between elements of the setting closure, freezing of the point to slide nuts, and freezing of the setting closure elements may block turnouts. Proper operation of turnouts improves the efficiency and safety of railway traffic.

Three methods of turnout heating are the most common in European countries:

a) electrical turnout heating by means of resistance heaters (power is mostly supplied from 3X400 [V] networks, 230 [V] networks or traction 15 [kV] $16\frac{2}{3}$ [Hz] systems via a voltage reducing transformer),
b) gas turnout heating,
c) water turnout heating of small railway facilities, mainly in Germany. The first system of this kind has been used at Boguszów station in Poland [2], [5].

2 Induction Heating of Railway Turnouts

PKP undertook initial testing of induction heating of turnouts (‘ior’ - an original Polish concept) in 1978/1979, involving five turnouts at Poronin station and 26

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 96–106, 2010.

turnouts of Tarnów West station. Insulation-coated heating bars were used. The bars were not in a galvanic contact with rails. The operating principle of the 'ior' consisted in heating a rail with eddy currents induced therein. A heating bar also heated to between +15°C and +20°C [1].

The heating bars were made of copper wrapped with Tarflon tape and placed inside a steel cover. They were supplied with a voltage of 3-3.3 [V] and frequency of 50 [Hz], while the current inside a bar was up to 350 [A].

The following types of heating bars were applied:

- 2.55 [m], power 750 [W], for UIC-49 rails
- 3.00 [m], power 900 [W], for UIC-60 rails.

2800 [VA] transformers powered all types of the heaters. At network frequencies of 50 [Hz], the heating bars vibrated and a human-audible acoustic wave occurred of a frequency double that of the supply voltage.

The induction nature of the loads of the then 'ior' devices applied to power supply systems necessitated the introduction of an additional capacitor to compensate the reactive power in order to improve the power factor $cos\varphi$ from approx. 0.5 to 0.85-0.9. Capacitors capable of correcting the reactive power by 4 [kVA] or more were used in a single turnout. The 'ior' equipment induced minor longitudinal voltages in rails. In normal operating conditions, compensation heaters reduced the voltage to several dozen millivolts, which did not interfere with track circuits. When several heaters broke at the same time, it could have risen to 0.3 [V], still not interfering with the operation of track circuits.

The scant materials concerning those 'ior' tests presented in [1] are the only discussions extant, available in archives of the then COBiRTK (Railway Engineering Research and Development Centre), today named IK (Railway Institute).

3 The Phenomenon of Induction Heating

The phenomenon of induction heating may be explicated using the example of a very flat plate (source of energy) across which a sinusoidally variable electric current of density J_z flows in the negative direction of the axis (Fig. 1a). The flow of current across the plate gives rise to an electromagnetic field both in and around the plate. If a large metal block - an energy receiver whose surface towards the side of the energy source is flat and parallel to the latter - is placed in the plate's vicinity (Fig. 1b), it will be affected by the electromagnetic field. The effect can be characterised with the aid of magnetic and electric field intensities on the block's surface. The intensity of the electric field $\underline{E}$ in the gap between the source and the receiver has two components: one, $\underline{E}_s$ is parallel to the z axis and surfaces of the source and the receiver, while the other is perpendicular to the source and the receiver, being directed from the former towards the latter (this component is not shown in Figure 1a).

The intensity of the magnetic field comprises only one component $\underline{H}$ along the y axis, whose value is constant, as opposed to $\underline{E}_s$ which changes its value and direction (Fig. 1a). The value of the electric field intensity component $\underline{E}_s$ is equal to $\underline{E}_{s0}$ on the receiver surface. The magnetic field component $\underline{H}$ is parallel to the receiver and the source, with its value on the surfaces designated by $\underline{H}_0$.

According to the right-hand rule, a portion of the electromagnetic field energy will be moved from the gap between the source and the receiver, where $\underline{E}_s$ is positive, towards the x axis in a direction perpendicular to $\underline{E}_s$ and $\underline{H}$ and will penetrate into the receiver. Both these magnitudes are generated by the sinusoidally variable current and will be sinusoidally variable as well. The electromagnetic wave radiating across the source towards the receiver also undergoes sinusoidal variations, with its two components $\underline{E}_s$ and $\underline{H}$ perpendicular to one another. The speed of wave radiation, ergo of the energy transport in the air filling the space between the source and the receiver, is equal to the speed of light $v = 3*10^5$ [km/s]. At the wave frequency $f = 50$ [Hz], its length is [3]:

$$\lambda = \frac{v}{f} = 6000[km] \tag{1}$$

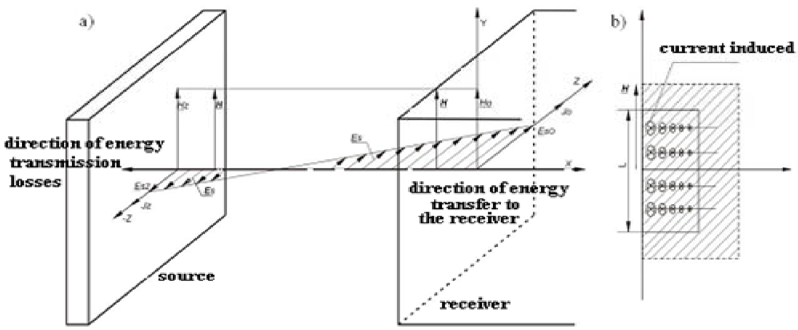


Fig. 1. a) Electromagnetic field between the source and the receiver: $\underline{H}$ – vector of magnetic field intensity, $\underline{E}_s$– static component of the electrical field intensity vector, $\underline{E}_{s0}$ – value of the static component of the electric field intensity vector on the receiver surface, $\underline{E}_{sZ}$ - value of the static component of the electric field intensity vector on the source surface [3], b) current induced in the receiver's plane [7].

When an electromagnetic wave λ long moves from the air to a conductor, its length reduces to a fraction of a metre and its speed to several metres a second. A wave penetrating a receiver will also be strongly damped and will discharge its energy near the surface. This is explained by the electric field triggering the flow of currents across the receiver since the latter is a conductor. These currents are referred to as eddy currents. They generate their own magnetic field, with a direction counter to $\underline{H}_0$ and weakening the latter. The weakening will be the greater, the further away from the surface. The frequency of these currents is equal to that of the penetrating wave.

The energy, in direct proportion to the electric intensity multiplied by the magnetic intensity and to the time elapsing since penetration of the receiver, is divided into an active and a reactive portion. The reactive portion is the part of electromagnetic energy which constantly shifts from its electric to its magnetic form (transferred from the electric to the magnetic field and vice versa) and does not contribute to the heat generation in the receiver. Only the active portion of the electromagnetic energy is transformed into heat [3].

By designating the portion of electric field intensity on the receiver surface as $\underline{E}_{s0}$ and the part of magnetic field intensity that causes the active power to be discharged in the receiver as $\underline{H}_0$ (as different from $\underline{E}_{s0}$ and $\underline{H}_0$ which characterize the overall power penetrating the receiver), the damping effect of the electromagnetic field can be defined:

$$H_x = H_0 e^{-\frac{x}{\delta}} \quad E_x = E_{s0} e^{-\frac{x}{\delta}} \tag{2}$$

where Hx and Ex are intensities at a distance of x from the receiver surface, δ – a constant called the depth of penetration [m], which is defined:

$$\delta = \sqrt{\frac{10^7 \rho}{4\pi^2 \mu f}} \approx 503\sqrt{\frac{\rho}{\mu f}} \tag{3}$$

where: μ - relative magnetic permeability of the receiver material, ρ – resistivity [Ω*m], f – frequency of the field variations [Hz].

The depth of penetration is a characteristic of induction heating that defines the distance *x* from a receiver surface at which the intensity of electric and magnetic field reduces *e* times compared to its value on the surface (Fig. 2).

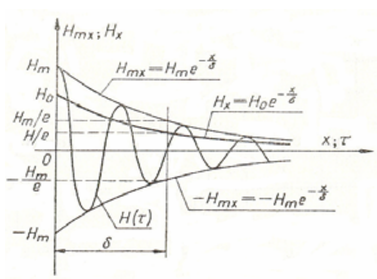


Fig. 2. Magnetic field penetration into an energy receiver H_m – amplitude of magnetic intensity on the receiver surface, H_{mx} – amplitude of magnetic intensity at a distance *x* away from the surface, H_x – root mean square of the magnetic field, $H(\tau)$ – momentary intensity of the magnetic field, δ - depth of penetration, τ - time, *e* – base of natural logarithms [3].

The greater the frequency of field variations and the magnetic permeability, the lower the penetration depth. Resistivity affects δ in the opposite manner. A bar parallel to the rail (rheostat) foot is the trigger of the magnetic field in the 'ior' system. The foregoing interpretation of distribution of the electric and magnetic fields must be undertaken in respect of the field of a round bar conducting the current. A greater frequency will affect the magnetic permeability of the receiver – which will diminish. Rail web and foot, whose magnetic permeability will have to be determined by laboratory testing, will be direct receivers in the 'ior' system.

Flows of eddy currents generate the heat energy in the receiver, causing its temperature to rise. To determine heating effects of the electromagnetic field, the distribution of the eddy currents or of the heating power, which is in direct proportion to the square of the current, must be determined. The power distribution will be referred to the individual elements of the receiver as the power discharged across each element will be different. In a system presented in Figure 1a, a receiver element at a distance of x from the surface has a length of dz, cross-section dS, and resistivity ρ. Let the eddy current I_x triggered by the electric field flows perpendicular to dS in the direction of the z axis and thus in the direction of the electric field E_x. The active power P_x discharged across such an element is defined as [3]:

$$P_x = I_x^2 \frac{pdz}{dS} \tag{4}$$

Multiplying the numerator and denominator of (4) by dS results in:

$$P_x = I_x^2 \frac{\rho dV}{(dS)^2} \tag{5}$$

where dV – volume of an element.

By dividing both sides of (5) by dV and determining the current density $J_x = I_x/dS$ a dependency defining the active power per unit of volume is produced:

$$\frac{P_x}{dV} = J_x^2 \rho = p_{vx} \tag{6}$$

The relation between Jx and magnetic intensity Hx in an element under consideration is defined as:

$$J_x = -\frac{\sqrt{2}}{\delta} H_x = -\frac{\sqrt{2}}{\delta} H_0 e^{-\frac{x}{\delta}} \tag{7}$$

(6) and (7) result in:

$$p_{vx} = \frac{2H_0^2}{\delta^2} e^{-\frac{2x}{\delta}} \tag{8}$$

where: pvx – active power of volume unit [W/m3], H0 [A/m], ρ [Ω*m], δ [m].

p_{vx} diminishes as a function of the distance x from the receiver's surface twice as rapidly as H_x, E_x, J_x (Fig. 3), where values of these magnitudes on the receiver's surface are indexed 0.

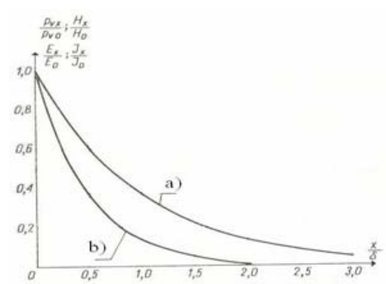


Fig. 3. Relative distribution of intensity of: a) magnetic field H_x/H_0, electric field E_x/E_0, current density $J_x/J_0\frac{J_x}{J_0}$, b) volumetric unit power p_{vx}/p_{v0} as a function of the relative distance from the plate surface x/δ [3]

(8) indicates that a mere 14% of power is discharged in a unit of volume at a distance of x = δ from the surface compared to the power released in the same unit of volume situated near the receiver's surface. (8) may also be employed to show that as much as 87% of power is discharged in the receiver's superficial stratum whose thickness is δ, that is, equal to the depth of penetration [3].

The active power penetrating inside a receiver across a unit of its surface, ps0 [W/m2], can also be defined on the basis of (8):

$$p_{s0} = H_0^2 \frac{\rho}{\delta} \tag{9}$$

The receiver is assumed to be very thick. As a result, the entire energy penetrating the receiver will be absorbed at some distance from its surface. Should the receiver be thin and the (primary) magnetic wave - upon reaching the surface opposite to the one it has penetrated – not be fully damped, it may be partly reflected (producing a secondary wave) and superimposed on the primary wave. This effect is analysed in detail in [3].

The foregoing equations indicate that empirical determination of penetration depth δ, associated with magnetic permeability μ (3), will define the area where eddy currents are generated in the rail, and thus the effectiveness of its heating. Current methods of turnout heating are based on the supply of a direct current and 50 Hz alternating current. The theoretical discussion implies the need for simulation and experimental verification of the effectiveness of turnouts heating by means of increased frequency currents and the assessment of energy effectiveness of this method in a broad range of supply voltage frequency variations. The dependencies presented above generally define the concept of induction heating and demonstrate its complexity. A definition of this phenomenon in the case of rail turnout heating requires detailed knowledge of the magnetic field distribution in a rail. The distribution of the magnetic field is closely related to rail's magnetic permeability μ (Fig. 2), which is dependent on a great number of factors (among others, contact stresses against the rail's superficial layer or the steel elasticity) and not amenable to simple definitions.

4 Complex Internal Structure of a Rail

Two main rail types are used on routes administered by Polish State Railways PKP PLK: UIC60 (currently marked 60E1) and S49 (currently 49E1). These designations were introduced when the EU legislation was implemented in Poland. Particular types are characterised by the weight of a running metre and cross-sectional dimensions. 49E1 (mass 49.39 kg/ running metre and cross-sectional surface 62.92 cm3) are used on routes with light rolling stock load. 60E1 rails (mass 60.21 kg/ running metre and cross-sectional surface 76.70 cm3) are used on heavily loaded routes where trains travel at speeds over 100 km/h.

Rail structures contain percentage admixtures of certain (metallic and non-metallic) elements that affect magnetic rail properties [4], [10].

As steel is heated and cooled and temperatures change in rail manufacturing processes, structural transformations occur. In the final production process, a rail is hot rolled at temperatures of 700 – 900°C [4]. The structure and material properties of the rail result from such processes.

Normalised dimensions of normal-gauge rail UIC-60 are illustrated in Figure 4. Based on these dimensions, a three-dimensional numerical rail model was generated for purposes of testing the distribution of induction and magnetic intensity. The knowledge of magnetic properties of a rail material is necessary to develop a method of induction turnout heating.

System's efficiency at optimum μ needs to be defined as a function of frequency when selecting the frequency of inductor power supply. Possible impact of the reflected wave, arising on the outside wall of the rail web, its counter-phase superimposition on the wave penetrating the receiver, and the resultant reduction of the system efficiency must be taken into account. These calculations should be executed at the stage of inductor design, and a correction of the adverse impact of the reflected wave may be attempted by means of frequency or phase of the supply voltage.

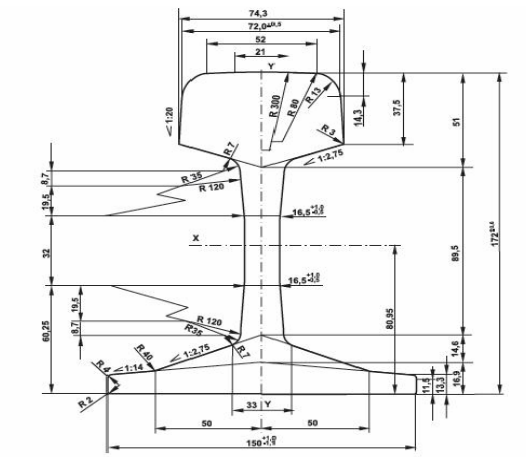


Fig. 4. Geometric dimensions of UIC-60 rail

Magnetic permeability μ of steel whose composition is very similar to that of the rail steel has been determined as μ=600 [8]. The source fails to describe the conditions of μ measurements, however. The analysis of *S*-60 rail cross-section detected a changed structure of the rail web edge cross-section and the rail head surface. These changes arose in the process of rolling and may substantially affect magnetic parameters of the rail, particularly in its superficial layer. To determine the rail structure changes and their effect on the efficiency of the inductor, samples of the web cross-section were cut and tested for magnetic permeability.

Material properties have been tested at the Silesian University of Technology in Katowice. The internal rail structure comprises several areas of varying magnetic properties (Fig. 5). The head and web edges acquire different material properties. Magnetic properties of the remaining rail sections (e.g. the core) are identical.

The 3D model was extended to include areas in respect of which separate prime magnetising curves were obtained, proof of structural differences arising at the time of production process rail rolling.

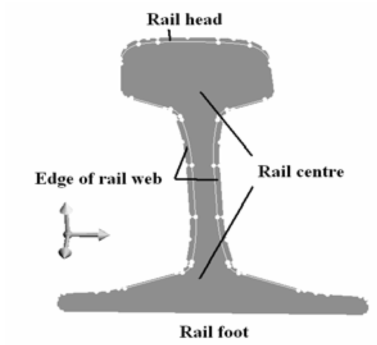


Fig. 5. Rail division into areas of diverse magnetic properties (prime magnetising curves)

The testing has demonstrated the occurrence of a limit frequency of 100 - 1000 Hz at which μ dramatically reduces. It can be preliminarily concluded that the range of induction heating operating frequency will not exceed 1000 Hz. Further testing of induction heating systems is therefore projected with regard to frequencies of 100 - 1000 Hz. Figure 6 shows magnetic permeability μ variations for a rail foot sample as dependent on variations of magnetic field frequency f and magnetic intensity $H = 0.5$ [A/m].

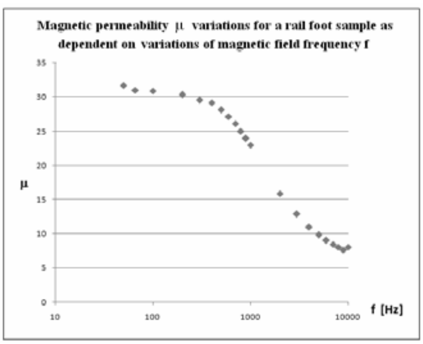


Fig. 6. Magnetic permeability μ variations for a rail foot sample vs. variations of magnetic field frequency f

Fig. 7 shows prime magnetising curves for the web edge, head, and core samples obtained at the time of laboratory testing. The curves shown below were used to define areas of separate magnetic properties in the presented model.

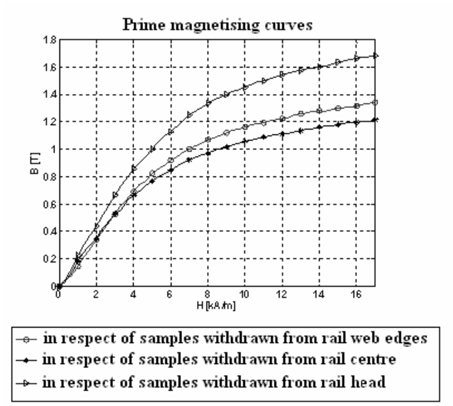


Fig. 7. Prime magnetising curves for web edge, head, and core sections

5 Three-Dimensional Model of UIC-60 Rail

A three-dimensional model of a rail helps to analyse magnetic and electric effects in the internal rail (material) structure under the influence of a magnetic field.

Figure 8 presents a numerical rail model executed in Flux 3D software and based on three-dimensional finite elements method in consideration of geometric rail dimensions (Fig. 4) and areas of varied magnetic properties (Fig. 5).

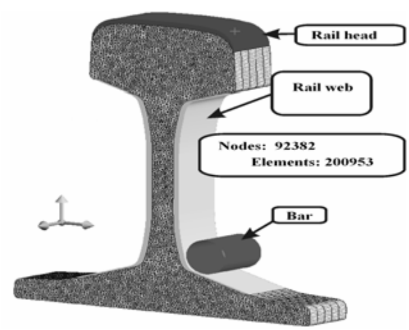


Fig. 8. Numerical model of a railway rail

6 Distribution of Magnetic Induction and Intensity across a Rail

A range of computer simulations were conducted on the basis of the numerical model. The calculations include the following simplifications: the magnetic hysteresis is ignored and density is assumed to be constant all across the bar cross-section. Figure 9 shows a sample distribution of absolute values of the magnetic induction at a current density of 11.14 A/mm^2.

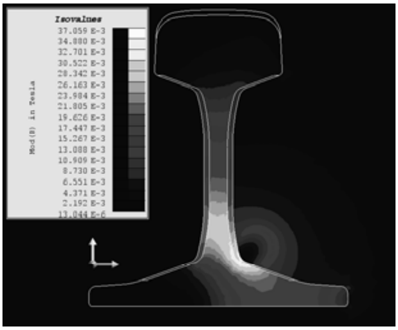


Fig. 9. Distribution of magnetic induction across a rail and in its environment

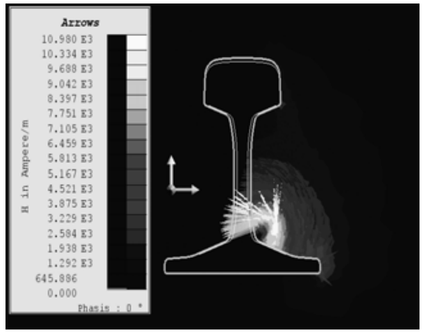


Fig. 10. Intensity of a magnetic field generated by a current across a bar

The design of the rail itself is not a magnetically closed circuit, the magnetic field is therefore very much dispersed in space (air) (Fig. 10). This weakened magnetic field may not suffice to generate eddy current in the rail, which will prevent the induction heating from being applied to railway turnouts. The maximum induction on the rail surface reached only 0.037 T. The intensity of a magnetic field generated by a current across a coil is shown in Figure 10.

7 Conclusions

When the ‘ior’ is introduced and designed, the following elements of a heating circuit must be analysed: the structure of the rail material, rail resistivity and magnetic permeability, skin effects, penetration depth of magnetic field into the rail structure, and release of active power in the area of the magnetic field affecting the rail [6]. These parameters depend, among other factors, on the magnetising current frequency.

Rail parameters: the resistivity, magnetic permeability or heat conductivity are not uniquely identified by rail manufacturers and require experimental determination (rails may be manufactured from different charges and subject to varied rolling, straightening, and possibly hardening technologies). It is therefore necessary to develop a numerical model of rail that is as precise as possible and preserves its structural complexity and properties.

The three-dimensional rail model that the authors have developed using Flux3D provides for an accurate interpretation of electric and magnetic phenomena [9] in the internal rail structure across which the eddy currents flow. The model may be employed, inter alia, to determine eddy and histeresis losses and penetration depth of the magnetic field into the rail structure (the parameters whose knowledge is necessary for the development of the ‘ior’ method).

References

1. Brodowski, D., Andrulonis, J.: Efektywność ogrzewania rozjazdów kolejowych, IK Warszawa (2000) (in Polish)
2. Brodowski, D., Andrulonis, J.: Ogrzewanie rozjazdów kolejowych. Problemy kolejnictwa zeszyt 135, IK, Warszawa (2002) (in Polish)

3. Gozdecki, T., Hering, M., Łobodziński, W.: Urządzenia elektroniczne. Elektroniczne urządzenia grzejne, Wydawnictwa Szkolne i Pedagogiczne, Warszawa (1979) (in Polish)
4. Grobelny, M.: Budowa, modernizacja, naprawa i remonty nawierzchni kolejowej – urządzenia i elementy, Rynek kolejowy (2009-03-09) (in Polish)
5. Kiraga, K., Szychta, E., Andrulonis, J.: Wybrane metody ogrzewania rozjazdów kolejowych – artykuł przeglądowy, Przegląd Elektrotechniczny, nr 2/2010 (in Polish) ISSN 0033-2097
6. Liwiński, W.: Nagrzewanie indukcyjne skrośne, Wydawnictwo Naukowo-Techniczne Warszawa rok (1968) (in Polish)
7. Sajdak, C., Samek, E.: Nagrzewanie indukcyjne. Podstawy teoretyczne i zastosowanie, Wydawnictwo "Śląsk" Katowice (1985) (in Polish)
8. Struktura materii, Przewodnik encyklopedyczny, Państwowe Wydawnictwo Naukowe, Warszawa (1980) (in Polish)
9. Szychta, E., Luft, M., Szychta, L.: Method of designing ZVS boost converter. In: Proceedings of the 13 International Power Electronics and Motion Control Conference, Poznań (2008)
10. Wielgosz, R.: Łączenie bezstykowe szyn kolejowych, Mechanika Czasopismo Techniczne, Wydawnictwo Politechniki Krakowskiej, 2-M/2009, Zeszyt 6, Rok 106 (in Polish)
11. Mikulski, J., Białoń, A.: Wpływ typu ogrzewania rozjazdów na zużycie energii elektrycznej. Przegląd Elektrotechniczny 9, 37–39 (2009)
12. Mikulski, J., Białoń, A.: Analiza zużycia energii w systemach elektrycznego ogrzewania rozjazdów. Technika Transportu Szynowego 1(2), 70–72 (2009)

Problems and Issues with Running the Cycle Traffic through the Roundabouts

Elżbieta Macioszek, Grzegorz Sierpiński, and Leszek Czapkowski

Silesian University of Technology, Department of Traffic Engineering,
Krasinskiego 8, 40-019 Katowice, Poland
{Elzbieta.Macioszek,Grzegorz.Sierpinski,
Leszek.Czapkowski}@polsl.pl

Abstract. The paper deals with the bicycle traffic issues on the roundabouts and their nearby areas. The bicycle traffic characteristics as well as pros and cons of a bicycle as a means of transportation with the current situation of cycling in Poland have been presented. The authors also presents the results of research on traffic safety of the cyclists on the traffic circles. In addition, the fields of knowledge lacking complete information on the solutions to bicycle traffic on the roundabouts are being pointed out.

Keywords: Bicycle traffic characteristics, roundabout, cycling in Poland.

1 Safety of the Cyclists on the Roundabouts

According to the research being carried out in different countries (e.g. [8]) the roundabouts seem to be much safer than any other types of intersections, especially for such road users as pedestrians and vehicle drivers. Based on this research, it has been estimated that a number of the road accidents involving pedestrians and vehicle drivers has dropped enormously after turning the perpendicular intersections into the roundabouts. Such decrease of the road incidents varies between 36% and 61% and depends on the country where research had been held, as shown below [12]:

- Austria – between 41% and 61%;
- Germany – 36%;
- Netherlands – 47%;
- the USA – 37%.

Also a reduction of the road incidents involving all kinds of casualties has been noticed. Such decrease varies between 25% and 87% and depends on the country where the research had been held, as shown below [12]:

- Austria – between 45% and 87%;
- France – between 57% and 78%;
- the UK – between 25% and 39%;
- the USA – 51%.

Opinions on the safety quality on roundabouts for the road users such as cyclists are split. Some believe that roundabouts are not any less safe to the cyclists than any

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 107–114, 2010.

other intersections whilst the others think differently. Studying research on the subject one may find results proving that in some cases, after turning the intersections into the roundabouts, a number of the road traffic accidents with the cyclists involved has increased.

According to the principles of the sustainable transport development, the cycle traffic shall be an evenly balanced alternative to the vehicle traffic. It is all about the transit rides in general and not only the recreational character of cycling. One of the most essential matters when it comes to introducing the cycle traffic to the towns and cities is providing the users with a sufficient level of the road safety.

The cycle traffic in Poland, in spite of lack of a complex network of the bicycle roads, has been systematically increasing for the last few years. The growth of people using a bicycle as a means of transportation does not go with the growth of a proper infrastructure for cyclists. There is lack of the sectioned bicycle paths, bicycle lanes, contra-flow bicycle lanes, short bicycle junctions, bicycle gates or locks on the three-leg and four-leg junctions. Cycle infrastructure in Polish cities and towns does not fulfill the current needs and therefore does not encourage people to go for a bicycle as a means of transportation.

In a general classification, the cycle traffic can use either roadways along with other vehicles or sectioned cycle tracks. The latter, according to the Journal of Laws [13], should be located in such way that would provide the cyclists with the road traffic safety.

A bicycle road (a bicycle path) is an independent road or a part of a road dedicated to the pedal cycles and marked with a "Road to be used by pedal cycles only" sign (a C-13 sign in Poland) [11]. According to the Vienna Convention on Road Traffic [18] it should be segregated structurally from different parts of the road such as roadway or sidewalk by the green belt or difference in the surface or height. The bicycle roads and paths can be either one-way or two-way.

Despite the increasing tendency of cycling in Poland the cycle traffic is still on a fairly low level in comparison to the countries such as Sweden, Netherlands, Belgium or Germany, which promote cycling and active lifestyle on a large scale. This is also due to lack of a bicycle road network of proper density and quality. In most cases cyclists in Poland use the same lane as any other vehicle users if there is no a bicycle path on the intersection or cross a street with pedestrians pushing their bicycles. In spite of a relatively rich collection of the abroad results of research on routing the bicycle paths on roundabouts as well as the road traffic safety issues regarding the pedestrians on those roundabouts, there is still shortage of the appropriate databases in Poland that would hold detailed data on the road traffic accidents when cyclists are involved. And such data should contain information about the place where an accident has occurred, available infrastructure and whether there is or there is not a bicycle path on this particular roundabout. In addition, there is also a huge inconsistency in all the collected solutions on the cycle traffic on roundabouts comparing to the countries where cycling and bicycles as means of transportation are in order.

Unfortunately, regardless whether it concerns Poland or any other country, the drivers generally do not fully respect those sharing the same space (non-sectioned lanes) but traveling much slower, which applies to the cyclists. Therefore, the most dangerous for the cyclists are level crossings and passages as well as all kinds of intersections.

The accidents on intersections with the cyclists involved are mainly caused by the fact that the vehicles traveling with different speeds change lanes and do all kinds of driving maneuvers including making changes in directions. The most dangerous to the cyclists are the multilane intersections and large junctions as they need to concentrate hard and pay constant attention to many different factors. This of course includes not only all behavioral and non-behavioral factors but also undertaking several important decisions in a very short period of time.

There are different ways of improving the safety quality of the cyclists in the intersection areas. This concerns both the infrastructural solutions and the traffic management on the intersections.

2 Pros and Cons of a Bicycle as a Mean of Transportation

The modern societies aim to popularize a bicycle as one of the main means of transportation. In Poland, as well as in many other countries it is being done by a program called “urban bicycle.” The bicycle as a means of transportation in everyday usage, especially in rides to and from school or work, has a number of advantages. It also has some disadvantages. The advantages of a bicycle as a mean of transportation are as following [6]:

- speed of movement. In the areas with a heavy traffic a bicycle is the fastest and most efficient form of transportation, especially on a distance of travel of 5 to 6 kilometers. The bicycle has such advantages of the individual means of transportation as the privacy and possibility of traveling from door-to-door;
- economics. The social costs of traveling by bicycle are much lower than the costs of traveling either by car or public transportation. These include the costs of building and maintaining the bicycle paths as well as the costs of manufacturing a bicycle itself and its current exploitation costs;
- energy efficiency;
- health values;
- minimal burden of the natural environment.

When it comes to disadvantages of a bicycle as a means of transportation one may find the following [6]:

- sensitivity and response to the weather conditions;
- a risk of the road accidents;
- dependence on the condition of the technical infrastructure. The full capability of usage of the bicycle depends on the density of the road networks and bicycle paths, condition of the road surface, parking capabilities and so on;
- a risk of the muggings and assaults, especially in lack of lighting and routing through the uninhabited and desolated areas.

3 Cycle Traffic in Poland

As it has been already mentioned, a growing number of bicycles in the road traffic in Poland did not gain full acceptance of the road users, especially vehicle drivers. This can also be seen in the road rules and regulations applying to the cyclists. According to the Traffic Code (Rules of the Road) there are two instances that apply to all road users [11]:

- A vehicle driver approaching a cycle crossing is obligated to pay special attention and give way to any bicycle being on that crossing.
- On a cycle crossing and directly ahead of it, a vehicle driver is forbidden to overtake any vehicle except for a crossing with traffic being controlled by authorized persons or traffic lights.

Since 2000, according to the Road Traffic Act 2000, drivers have been granted no obligation to give way to cyclists riding bicycles on a path that runs through a road that a driver enters or turns to. Thus, the right of way to cyclists has been ruled out [19].

Cycling in Poland, following the worldwide trend, has stopped having only a recreational character and is slowly becoming a more popular means of transportation, especially when traveling to school or work. There were approximately 20,000 km of the traced bicycle routs in Poland in 2008 [16]. The Ministry of Infrastructure has announced the continuation in 2010 of an improvement program of building and developing the shared zones (the footpaths combined with the bicycle paths). A campaign called "Switch the thinking on" ("Włącz myślenie") and promoting safety on roads has also been announced. The main aim of the campaign is to bring the cyclists, pedestrians and other road users at risk to attention of drivers ([10]). Only in few Polish cities –including Cracow, Warsaw and Danzig– development of the bicycle paths and roads is fully advanced and all kinds of actions promoting a bicycle as a mean of transportation and cycling in general take place. Unfortunately the majority of Polish cities have no such fortune. In many places the situation of the cyclists is still bad and, what is more, the bicycle paths do not exist at all.

Routing the bicycle paths in urban parks and open spaces –that is in some distance of the streets with heavy traffic– is not difficult and it is basically doable to find a suitable space for such aim. A problem appears in cities.

Polish experiences show that extracting the bicycle paths off the streets is difficult. Two main causes of such status are the road and traffic management authorities, and rules and regulations regarding planning. The first applies when the local officials treat such solutions as dangerous. The latter applies when the rules and regulations do not allow routing the bicycle paths off the street in case when the street side parking is allowed. This is due to lack of possibility of marking such correctly. However, it appears in the majority of European cities and is successfully applied to [4].

In the areas of intersections the vehicles and bicycles use the same lane. Such solution is commonly used in Poland except for a few new and innovative projects, as in Warsaw on Emilia Plater Street, where the bicycle path linking the city center with the campus of the Warsaw University of Technology runs through the intersections. One needs to bear in mind that the increasing intensity of the bicycle flow and the growing popularity of a bicycle as a means of transportation will cause a need to provide the cyclists with a sufficient level of safety on those intersections. This, on the other hand, may lead to the development of infrastructure.

As it has been indicated, in [5] Poland it is of the highest risk when it comes to cyclists and road traffic safety. Each year police statistics read a few thousand of the road collisions involving the cyclists. An average number of 500 bicycle users gets annually killed in all kind of accidents and over 5,000 gets seriously wounded. It is a huge problem in large cities and small towns. Most of the road accidents take place in the cities but those of the heavier casualties happen in the smaller towns and villages.

Each year the Department for Road Traffic of the Police Headquarter issues the yearly reports regarding –among others– the road traffic safety of cyclists. This data includes only the incidents caused by the cyclists involved. It does not contain any information on what kind of accidents they are and how they occur. Moreover, it does not include any information on the accidents caused by the other road users and resulting with the casualties among the cyclists involved. However, this happens to be most common cause of all incidents occurring on the Polish roads. The police reports only read a general number of the road incidents that involve the cyclists as well as a general number of casualties of all kinds. Lacking the precise data makes it impossible to set a sufficient level of the road traffic safety for cyclists. All these issues require a wide knowledge of the field and additional research and analyses in the cycle traffic solutions.

4 Results of Abroad Research on Traffic Safety of the Cyclists on Roundabouts

It is impossible to carry out the comparing analyses on different solutions regarding running the cycle traffic on roundabouts in Poland. It is also impossible to indicate which method of running the cycle traffic on roundabouts gives the cyclists the proper safety conditions in terms of the safety quality. This is due to lack of the variety of cycle infrastructure in Poland. Therefore, abroad research on the subject has been studied and the results are being presented below:

- comparing roundabouts to other types of intersections, one may notice that the curvature of the routes for vehicles driving through on the roundabouts lowers the speed of ride. Such decrease of velocity makes safety of all traffic users running through the roundabout on the same lanes to increase;
- separating the bicycle paths on the roundabouts is essential when the road traffic volume is at least on a level of 8,000 vehicles/h and the cycle traffic volume is to be "considerable" [14];
- in Belgium, for the roundabouts smaller than 22m in radius the bicycle paths shall not be separated from the roundabout envelop by the road markings. The velocity of vehicles and cyclists passing through such roundabouts are very similar and the drivers do not tend to overtake the cyclists. In this case the bicycle path needs to end 15m before a "Give way to traffic on major road" sign (an A-7 sign in Poland). For the roundabouts larger than 22m in radius the bicycle paths need to be separated from the roundabout envelop by the road markings. Bigger dimensions of the roundabouts allow the vehicles passing through to run with higher velocities and the drivers tend to overtake

the cyclists. This definitely lowers safety of the cyclists riding through such roundabouts [17];

- research held in Belgium in 2007 proved that the intersections transformed into roundabouts had caused the growth of casualties and fatalities among the cyclists being involved in the road accidents on the roundabouts. The average number of casualties has grown up to 27% whilst the average number of fatalities has grown up to 41% - 46% [15]. Comparing the results, when it comes to the growth of safety of the roundabout users, one may notice that in case of the cyclists these are surprisingly bad. The decrease of safety of the cyclists after transforming intersections into roundabouts can be also found in other research, as in [9];
- the number of the road accidents on the roundabouts differs and depends on their location. In case of the roundabouts located in the urban areas, a total number of the road accidents with the cyclists involved grows averagely up to 48% in comparison to the accidents occurring on different types of intersections. Moreover, the growth of fatalities among the cyclists being involved in the road accidents can be estimated on a level of 77%. Research has also proven that, in the open space, the intersections with traffic lights after being transformed into roundabouts are less safe to the cyclists than other types of intersections transformed into roundabouts [15].
- small one-line roundabouts are safer for cyclists than two-lane and multilane roundabouts [3];
- Dutch research proves that the safest way of running the cycle traffic on the roundabouts is to route the separated bicycle paths on the outer side of the roundabouts. Moreover, the cyclists must give way to the oncoming traffic and the drivers receive the right of way [14];
- when it comes to the traffic safety it is a bad solution to route the cycle paths adjacent to the roundabout envelope on the outer side and to grant the cyclists the priority by giving them the right of way [1];
- the roundabouts with the lighter traffic volumes (less than 10,000 vehicles/24h and less than 1,000 cyclists/24h) are safer than the roundabouts with the heavier traffic volumes (values larger than these indicated above) [2];
- cyclists do not follow the traffic rules and regulations when entering and leaving the roundabouts. Research held in Holland proves that 13% of the cyclists ride their bicycles in the opposite direction than it is allowed to. Moreover, over 40% of the cyclists do not give way when entering the roundabouts [8];
- research [7] on the relationship between the yearly number of the road accidents involving the cyclists and the geometry of roundabouts, a period of time since the roundabout have been in use, as well as the traffic volumes of pedestrians and cyclists, and the velocity of vehicles had been held. Using the regression analysis it has been established that the number of the road accidents involving the cyclists per year depends on the traffic volumes of vehicles and cyclists. The bigger the traffic volume is the bigger probability of the collision to occur. It has been also established that the probability of collision with the cyclists involved grows with the period of time since the

roundabout have been in use. Also a certain relation between the yearly number of the road accidents and the velocity of vehicles has been noticed;
- based on information collected in 2006 from the Danish hospitals, it has been established that only 3% of all the road accidents with the cyclists being involved are reported to the Police [7].

These conclusions of research held abroad one would need to revise intensely and analyze acutely before similar infrastructure solutions be applied to Poland.

5 Conclusions

Cyclists as well as drivers are the equal users of any transportation system. Because of much smaller number of the cycle travels in Poland than in the countries of Western Europe, one does not find a large variety of solutions relating to the cycle traffic on roundabouts. Popularization of a bicycle as a means of transportation in every day usage is possible and according to the sustainable transport development policy is even necessary. However, it requires undertaking more drastic measures that will lead to the development of the bicycle infrastructure of all kinds. Undertaking such actions one needs to take the abroad experiences into consideration. Each time the choice of which solution to apply should be made independently and according to the location of the roundabout and all external factors having influence on how the traffic flows are forming.

Furthermore, riding a bicycle on the roundabout with the poor cycle infrastructure makes the bicycle users feel uncomfortable and less safely. A decision about using a bicycle as a means of transportation needs to be made after taking many different aspects into consideration. These factors include mostly the distance and time of travel, a degree of difficulty of the route, comfort and safety of the trip, noise and possibility of leaving a bicycle in a safe place. There are much more issues to think of but these are the most common. And the more concerns a cyclist has the more likely a bicycle will not be chosen as a means of transportation.

References

1. Brilon, W.: Roundabouts: A State of the Art in Germany. In: National Roundabout Conference, Vail, Colorado (2005)
2. Brude, U., Larsson, J.: What Roundabout Design Provides the Highest Possible Safety? Nordic Road and Transport Research, pp. 17–21 (2000)
3. Brude, U., Larsson, J.: The Safety of Cyclists at Roundabouts: a Comparison Between Swedish, Danish and Dutch Results. Swedish National Road and Transport Research Institute (1996)
4. Brzeziński, A., Jesionkiewicz–Brzezińska, K.: Rowerowa Europa. Przykłady Rozwiązań Służących Rozwojowi Ruchu Rowerowego. In: IV Konferencja naukowo-techniczna Miasto i Transport 2010. Obsługa Komunikacyjna Centrum Miasta, Warsaw (2010)
5. Buczyński, A., Hyła, M., Lustofin, B., Kopta, T., Obara, G., Rolla, M.: Zdarzenia Drogowe z Udziałem Rowerzystów 2006-2008. Studium/GDDKiA, Warsaw/Cracow (2009)
6. Gaca, S., Suchorzewski, W., Tracz, M.: Inżynieria Ruchu. Teoria i Praktyka, WKŁ (2008)

7. Hels, T., Orozova-Bekkevold, I.: The Effect of Roundabout Design Features on Cyclist Accident Rate. Accident Analysis and Prevention 39, 300–307 (2007)
8. Hyden, C., Varhelyi, A.: The Effects on Safety, Time Consumption and Environment of Large Scale Use of Roundabouts in an Urban Area: a Case Study. Accident Analysis and Prevention 32, 11–23 (2000)
9. Moller, M., Hels, T.: Cyclists' Perception of Risk in Roundabouts. Accident Analysis and Prevention 40, 1055–1062 (2008)
10. Pałys, E.: Bezpieczeństwo Niechronionych Uczestników Ruchu. In: Polskie Drogi. Bezpieczeństwo i Inżynieria Ruchu Drogowego, pp. 78–80 (2009)
11. Prawo o Ruchu Drogowym. Tekst Ujednolicony. Stan prawny na 22. 09 (2009)
12. Robinson, B., Rodegerdts, L., Scarborough, W., Kittelson, W.: Roundabouts: An Informational Guide. Federal Highway Administration (2000)
13. Rozporządzenie Ministra Transportu i Gospodarki Morskiej z dnia 2 marca, r., w Sprawie Warunków Technicznych Jakim Powinny Odpowiadać Drogi Publiczne i Ich Usytuowanie. Dz.U.43 poz. 430 (1999)
14. Schoon, C., Minnen, J.: The Safety of Roundabouts in the Netherlands. Traffic Engineering & Control 35(3), 142–148 (1994)
15. Stijn, D., Nuyts, E., Wets, G.: The Effects of Roundabouts on Traffic Safety for Bicyclists: An Observational Study. Accident Analysis and Prevention 40, 518–526 (2008)
16. Sulmicki, M.: Turystyka Rowerowa: Fakty i Mity. Zielone Światło. Biuletyn Centrum Zrównoważonego Transportu 14 (2008)
17. The Safety of Roundabouts and Traffic Lights in Belgium. In: National Roundabout Conference. Region Wallonne (2005)
18. Vienna Convention on Road Traffic, `http://www.unece.org/`
19. Zmiany w Ustawie: Prawo o Ruchu Drogowym (2000), `http://www.kodeksdrogowy.com.pl`

Methods of Modeling the Bicycle Traffic Flows on the Roundabouts

Elżbieta Macioszek, Grzegorz Sierpiński, and Leszek Czapkowski

Silesian University of Technology, Department of Traffic Engineering,
Krasinskiego 8, 40-019 Katowice, Poland
{Elzbieta.Macioszek,Grzegorz.Sierpinski,
Leszek.Czapkowski}@polsl.pl

Abstract. The paper deals with the bicycle traffic issues on the roundabouts and their nearby areas. The fundamental elements of traffic management and infrastructure used in traffic regulation on the roundabouts have been presented. The authors present also the examples of typical settings of the bicycle paths. Amongst the conventional solutions some interesting ones from abroad, from the Netherlands in particular, which grant a huge level of traffic safety while crossing a roundabout, have also been introduced.

Keywords: bicycle traffic, bicycle path, roundabout, traffic safety.

1 Introduction

For many reasons the modern roundabouts are being considered much safer than other types of the road intersections. The main effect on the safety of such junctions is due to a low velocity of the vehicles passing through as well as much lower number of the points of collision if comparing these to the other types of intersections. On the two-lane roundabouts, as on the contrary to the one-lane roundabouts, one will notice that there are some additional points of collision when changing lanes on the roundabout envelopes as well as when exiting the flow and the roundabout envelope. This lowers the road safety of the cyclists on such type of the roundabouts in comparison to the one-lane roundabouts.

There are different ways of improving the safety quality of the cyclists in the intersection areas. This concerns both the infrastructural solutions and the traffic management on the intersections. The fundamental elements of traffic management and infrastructure used in traffic regulation on the roundabouts have been presented below.

2 Methods of Routing Bicycle Paths on Roundabouts

As in case of other types of intersections it is possible to run the cycle traffic on the roundabouts in many different ways. The number of points of collision between the vehicles and cyclists on a roundabout depends on a location of a bicycle path or lack thereof. Hereafter some typical solutions applied to Poland and those one may find only abroad are being presented.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 115–124, 2010.

2.1 Non- sectioned Cycle Paths on Roundabouts

Non- sectioned bicycle paths on roundabouts are the most common solution. The cyclists and the drivers on the roundabouts are treated equally. In case when there are non- sectioned bicycle paths the cyclists need to use the same lanes as the drivers. Therefore, all traffic rules apply to them as much as to the other road users. The cyclists are considered as vehicles and let be ridden around the roundabout the same way like the drivers would have to do. Thus the vehicle and bicycle flows are mixed together, which creates some dangerous situations on the road (Fig. 1a, 1b). As it has been mentioned in some works [2], [3] the number of accidents where the cyclists are involved may –in such case– be greater than for other road users. This is due to both, the larger differences in velocities of the motorized users and cyclists as well as poor visibility of the cyclists. A number of these cross the roundabouts as the pedestrians do pushing their bicycles, especially if there is no bicycle path.

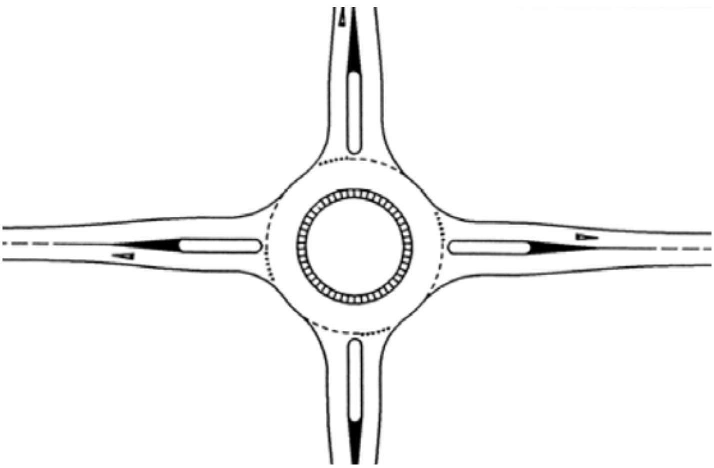


Fig. 1a. Roundabout with mixed traffic. Source: [4].

Fig. 1b. Roundabout with mixed traffic. Source: own work.

2.2 Sectioned Cycle Paths Adjacent to an Envelope with the Priority to Cyclists

Sectioned bicycle paths adjacent to a roundabout envelope with its inflow and outflow junctions provide the bicycle traffic with a spatial segregation from the vehicle traffic (Fig. 2a, 2b). The cyclists and vehicles have their own facilities, although there is no any physical barrier between them. In most cases the bicycle paths differ from the traffic lanes by the color and sometimes also by the surface structure. The cyclists receive priority as the drivers have to give way. Thus, the right of way to the oncoming bicycle traffic applies. However, according to research [1] such solution does not provide the cyclists with a sufficient level of safety and leads to many potentially dangerous situations. That includes as following:

- a risk of collision between the oncoming traffic and a cyclist on a cycle lane when a vehicle approaches the roundabout and enters the roundabout envelope;

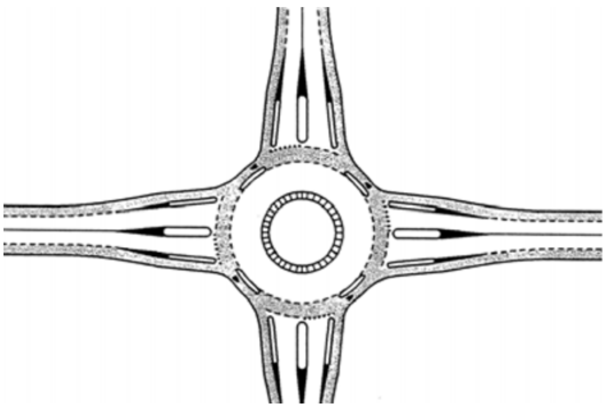


Fig. 2a. Roundabout with adjacent bicycle lanes. Source: [4].

Fig. 2b. Roundabout with adjacent bicycle lanes. Source: [5].

- a risk of collision when vehicle enters the roundabout envelope and a driver may not see the oncoming cyclist, especially when both use the same inflow junction;
- a risk of collision when drivers of the vehicles already circulating around the roundabout may not see a cyclist in a rear mirror (blind area), especially when the latter uses the cycle lane adjacent to the roundabout envelope.

2.3 Sectioned Cycle Paths with the Priority to Cyclists

Cyclists on a roundabout use the bicycle paths sectioned from a roundabout envelope by a central reservation either in form of an island elevated above the level of the roundabout or in form of a green belt. Such solutions ensure a sufficient level of safety to all cycle traffic as the cyclists crossing the approaching and exiting vehicle traffic receive priority. Hence the vehicle drivers have to give way and the right of way to the oncoming bicycles applies (Fig. 3a, 3b).

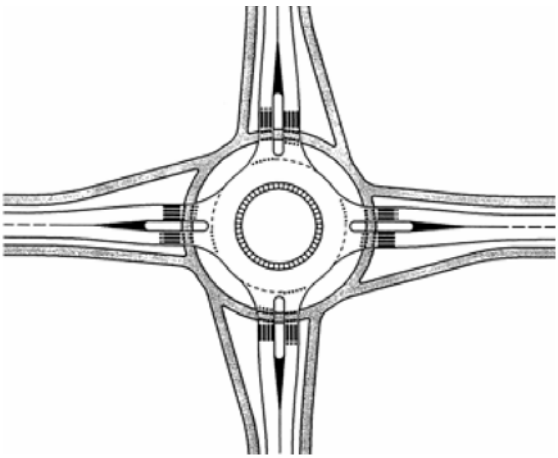


Fig. 3a. Roundabout with sectioned bicycle lanes. Priority given to the cyclists. Source: [4].

Fig. 3b. Roundabout with sectioned bicycle lanes. Priority given to the cyclists. Source: [7].

The cyclists ride on a sectioned path when circulating the roundabout. Safety of the cycle traffic depends on the mutual awareness of the presence of the bicycles and motorized vehicles in traffic.

2.4 Sectioned Bicycle Paths without the Priority to Cyclists

Cyclists on a roundabout use the bicycle paths sectioned from a roundabout envelope by a central reservation. The cyclists crossing the approaching and exiting vehicle traffic have to give way. Thus a priority rule to the oncoming vehicles applies (Fig. 4a, 4b).

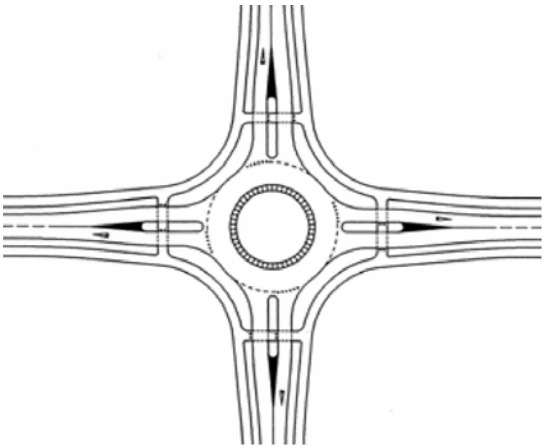


Fig. 4a. Roundabout with sectioned bicycle lanes. No priority given to the cyclists. Source: [4].

Fig. 4b. Roundabout with sectioned bicycle lanes. No priority given to the cyclists. Source: [6].

2.5 Footbridge for Pedestrians and Cyclists Elevated over an Intersection

A footbridge for pedestrians and cyclists in a shape of circle elevated over an intersection is a solution that eliminates all kinds of collisions between the vehicle traffic, pedestrian traffic and cycle traffic (Fig. 5a, 5b). There are four paths with a 3% slope each leading to the footbridge, which also allows skaters and handicapped people to use the passage without difficulties. Such multilevel solutions one can find in Holland, in Eindhoven for that instance. They can be applied over any type of intersection, not only over the roundabouts.

Fig. 5a. Footbridge for pedestrians and cyclists elevated over an intersection. Source: [7].

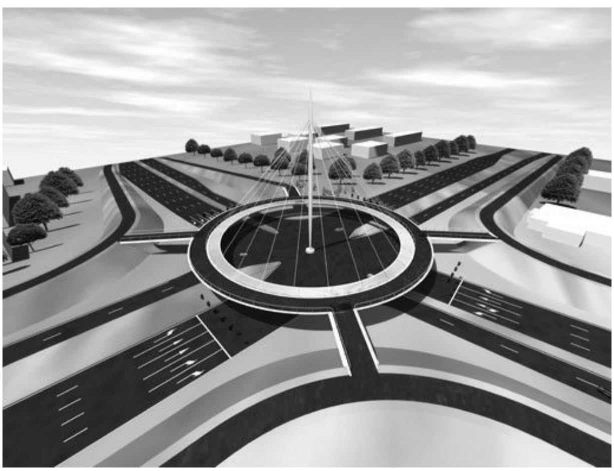


Fig. 5b. Footbridge for pedestrians and cyclists elevated over an intersection. Source: [8].

2.6 Sectioned Bicycle Paths under an Elevated Roundabout

Another solution eliminating all kinds of collisions between the vehicle traffic, pedestrians and cyclists is to build an elevated roundabout for all vehicle traffic and to design the cycle paths under the roundabout (Fig. 6a, 6b). This is a very interesting and yet a very simple solution in terms of the traffic organization and it provides the cyclists with a high level of safety. Differing from the majority of projects developed and solutions applied, this time it is the roundabout that has been elevated over the ground level. This makes the usage of the passage much easier as the pedestrians and cyclists do not need to use deep slopes or cross the intersection via dark, narrow and sometimes even unsafe underpasses. Such solutions one can find mainly in Holland, in the Eindhoven area for that instance.

Fig. 6a. Sectioned bicycle paths under an elevated roundabout. Source: [4], [8].

Fig. 6b. Sectioned bicycle paths under an elevated roundabout. Source: [4], [8].

2.7 Roundabout with an Open-Up Bicycle Tunnel

The vehicle and bicycle flows run at two different levels (Fig. 7a, 7b). The vehicle flow runs on the ground level and the bicycle flow runs underneath the roundabout in a bicycle tunnel. For the social safety reasons a decision to open up the tunnel in the middle of the roundabout has been made. These solutions are applied to the open air areas in case of a significant volume of the cycle traffic flows in an intersection zone as well as frequent collisions involving the cyclists. Such solution can be found in Holland (in the Dronten area for that instant).

Fig. 7a. Roundabout with an open-up bicycle tunnel. Source: [7], [8].

Fig. 7b. Roundabout with an open-up bicycle tunnel. Source: [7], [8].

2.8 Roundabout with a Bicycle Path in the Middle of the Road

One of the most interesting although not that common solutions is an example of a bicycle roundabout in Lelystad, Holland. On this particular roundabout a lane for cyclists is to be found not on the outside, but in the middle of a traffic lane (Fig. 8a). As a result of such layout the drivers cannot overtake the cyclists on the roundabout envelope. They can only drive behind or in front of the cyclists, which effectively improves safety of the cyclists on this roundabout. The cyclists enter the roundabout via a bicycle path that makes one of the leg junctions. This pilot project came to existence in 2006 and as of now is still experimental.

Fig. 8a. Roundabout with a bicycle path in the middle of the road. Source: [7], [8].

When approaching the roundabout one can find special traffic signs that inform the drivers about the presence of cyclists in the middle of the road (Fig. 8b). Along with the emblem of a cyclist the caption reads: "Caution: bicycle roundabout. All vehicles ahead of or behind, NOT beside a cyclist."

Fig. 8b. Roundabout with a bicycle path in the middle of the road. Source: [7], [8].

3 Conclusions

All the solutions presented in the article have their advantages and disadvantages. Each time the choice of which solution to apply should be made independently and according to the location of the roundabout and all external factors having influence on how the traffic flows are forming.

One should also bear in mind that apart from the optimal traffic organization on the roundabouts, there are also the safety issues to be concerned about. Safety of a cyclist in the late hours depends mainly on the proper illumination of the roundabouts and all side streets in the area of those roundabouts as well as their visibility. A good visibility can be provided by wearing light-colored or fluorescent clothing and/or reflective accessories which will help other road users to see a cyclist in the dark or in poor light.

Riding a bicycle on the roundabout with the poor cycle infrastructure makes the bicycle users feel uncomfortable and less safely. A decision about using a bicycle as a means of transportation needs to be made after taking many different aspects into consideration. These factors include mostly the distance and time of travel, a degree of difficulty of the route, comfort and safety of the trip, noise and the possibility of leaving a bicycle in a safe place. There are much more issues to consider, but these are the most common. And the more concerns a cyclist has the more likely a bicycle will not be chosen as a means of transportation.

References

1. Brilon, W.: Roundabouts: A State of the Art in Germany. In: National Roundabout Conference, Vail, Colorado (2005)
2. Brown, M.: The Design of Roundabouts. State of the Art Review. Transport Research Laboratory, London (1995)
3. Roundabouts: An Informational Guide. Report No. FHWL-RD-00-67, Washington, DC (2000)
4. Stijn, D., Geert, W.: Traffic Safety Effects of Roundabouts: a Review with Emphasis on Bicyclist's Safety. In: 18th ICTCT Workshop in Helsinki, Finland (2005)
5. Motorcycles travels, `http://bernys.ovh.org/`
6. Bicycles Bialystok, `http://www.rowerowy.bialystok.pl/`
7. Fietsberaad: fietsberaad.nl
8. Bicycle Wroclaw Portal, `http://rowery.eko.org.pl/`

Standardization and Interoperability Problems of European Electronic Tolling Service (EETS)

Gabriel Nowacki[1], Izabella Mitraszewska[1], Tomasz Kamiński[1], Włodzimierz Potapczuk[2], and Thomas Kallweit[3]

[1] Management and Transport Telematics, Motor Transport Institute, 80 Jagiellońska Str, 03301 Warsaw
gabriel.nowacki@its.waw.pl
[2] AUTOGUARD S.A., 27 Omulewska Street, 04128 Warsaw
w.potapczuk@autoguard.pl
[3] FELA Management AG., Basadingerstasse 18, CH-8253 Diessenhofen
thomas.kallweit@fela.ch

Abstract. The paper refers to some standardization and interoperability problems of the European Electronic Toll Service (EETS) implementation in European Union. The existing EETS systems in the European Union member states are not interoperable due to many differences among them. European Commission has taken bold steps to address that issue. The first one was the 2004/52/EC Directive on the interoperability in the Community. The second one was the decision to launch Europe's own Galileo system. The third was the EC decision from 6th October 2009, based on Research Charging Interoperability (RCI) and the Common Electronic Fee Collection System for a Road Tolling European Service (CESARE) projects. Furthermore, the Motor Transport Institute researches, concerning the mentioned matters have been presented too.

Keywords: Research Charging Interoperability (RCI), Common Electronic Fee Collection System for a Road Tolling European Service (CESARE).

1 Introduction

There are two different types of the European Electronic Tolling Service (EETS): the Dedicated Short Range Communication (DSRC) and the GPS/GSM based systems.

The electronic toll collection systems in the European Union member states are not interoperable due to the differences in charging concepts, technology standards, classification and tariff structure, legal and institutional backgrounds. European Commission has taken bold steps to address that issue. The first one was the 2004/52/EC Directive of the European Parliament and of the Council of 29th April 2004 on the interoperability of electronic road toll systems in the Community [3]. The second was Commission Decision of 6 October 2009 on the definition of the European Electronic Toll Service and its technical elements.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 125–132, 2010.

The requirements of that directive will be implemented in Poland based on the Act from 7 of November 2008 and some other acts [6]. It stressed that toll collecting charge institutions should be able to carry out electronic toll transactions from 1st July 2011.

Interoperability of road charging solutions is a long-term objective of the EC and as mentioned earlier, the 2004/52/EC Directive of the European Parliament and Council on the interoperability of the electronic road toll systems in the Community was adopted in April 2004. The new road charging service that is interoperable throughout Europe on the basis of one or more of the mentioned technologies is called the European Electronic Tolling Service (EETS).

2 The Technical and Economical Aspects of EETS

According to data presented by EFKON AG [7], the implementation costs of Electronic Toll Collection System based on GPS/GSM technology are a little more (about 20 %) than those of the Dedicated Short range System (DSRC) at the beginning of implementation, including roadways below 1000 km and the assumption of 300 000 on-board units (OBU’s).

The costs are the same for DSRC and GPS/GSM systems when the roadways total length is 1000 km. Furthermore, for those above 1000 km, the costs of GPS/GSM based system are getting significantly lower (up to 60 % with total number of roadways – 3000 km), while the costs of DSRC system are decreased by only a few percent.

For road networks of over1000 km length, European Electronic Toll Service using GPS/GSM technology is more cost-efficient. Charging secondary roads/extending network with using GPS/GSM technology is the choice of implementation.

The profit from Electronic Toll Collection System (DSRC) in Czech Republic implementation was 213 million Euros in 2007 and 236 millions Euros in 2008. There were 357 000 registered OBU`s in 2008, and 380 000 OBU`s at the beginning of 2009.

The daily profit from using DSRC system in Czech Republic is 740 000 Euros. Based on analyzes it is known that profit of operating system will be 2,5 billion Euros for 10 years [1].

The implementation cost of Toll Collect System in Germany was about 1.24 billion Euros. Yearly profit from the system is 3.5 billion Euros – 4.022 billion Euros in 2009 (official date – Toll Collect GmbH/April 9-th 2010).

One of the reasons to introduce Toll Collect in Germany was a problem of efficient checking the trucks routing, especially as far as invaders and mistakes in fee calculations are concerned. The monitoring data: who, when and why goes this way in DSRC system and depends on many persons and more time, which increases the cost of the system operation.

Taking into consideration problems of microwave propagation, especially in urban and mountain areas, Czech Republic Government signed new contract with Kapsch in 2008 to implement hybrid system, which includes DSRC technology in actual roadways (972 km) and new GPS/GSM technology on new motorways and expressways.

Table 1. Comparative study of Electronic Toll Collection Systems in Europe

Characterization	Austria [4]	Czech [1]	Germany[1]
Introduction date	01.01.2004	01.01.2007	01.01.2005
Admissible weight (2010)	> 3,5 ton	> 3,5 ton	> 12 ton
System cost (Government contracts)	750 M€[2]	780 M€[3]	1240 M€[4]
Technology	DSRC	DSRC	GPS/GSM
Average charge	0,26 €	0,15 €	0,17 €
Budget revenues (2008)	1,026 B€	236 M€	3,5 B€
Operational and control costs	12 %	10 %	11,2 %

3 Road Charging Interoperability (RCI) Project

Within the framework of EETS researches the three-year (2005 – 2008) Road Charging Interoperability (RCI) project, which is partially funded by the DG Energy and Transport of the European Commission, was developed by Consortium currently consisting of 27 partners, including toll operators, suppliers, truck makers, representatives of both the DSRC and the GNSS[5] communities, and some specialist companies providing expertise on the relevant research issues [5].

RCI Project was implemented and tested the framework in field trials at six following sites: Austria (ASFINAG), Germany (TOLL COLLECT), Italy (TELEPASS), France (TIS PL), Spain (VIA-T), Switzerland (LSVA).

[1] According to German Ministry of Transport, Construction and Urban Development: www.bmvbs.de

[2] The company ASFINAG, responsible for the introduction of the EFC system, decided finally for the Italian motorway operator Autostrade S.p.a. as the best bidder in the evaluation process. Autostrade, respectively its Austrian subsidiary EUROPPASS signed a contract. The contract has an equivalent value of 750 million Euro. The implementation of the system took 18 months as it was requested in the tender. Source: *PIARC Seminar on Road Pricing with emphasis on Financing , Regulation and Equity.* Cancun, Mexico, 2005, April 11-13 1/8.

[3] The Transportation Ministry announced in early October (2005) that the system has cost 780 Million Euros for the Czech Republic, and was promised to give return on investment by 2010.

[4] Based on: www.bmvbs.de

[5] GNSS (Global Navigation Satellite System) is the standard generic term for satellite navigation systems that provide autonomous geo-spatial positioning with global coverage. GNSS-1 is the first generation system and the combination of GPS and GLONASS. GNSS-2 is the second generation of systems that independently provides a full civilian satellite navigation system, exemplified by the European Galileo positioning system.

RCI defined itself a high-level architecture for interoperability that is based upon work of the CEN and ISO standardization committees and the ASECAP[6] tolling operators' and Member States' Stockholm Group role model (CESARE III). There are main actors in mentioned model: the Toll Charger, the Toll Service Provider and the Service User (Fig. 1).

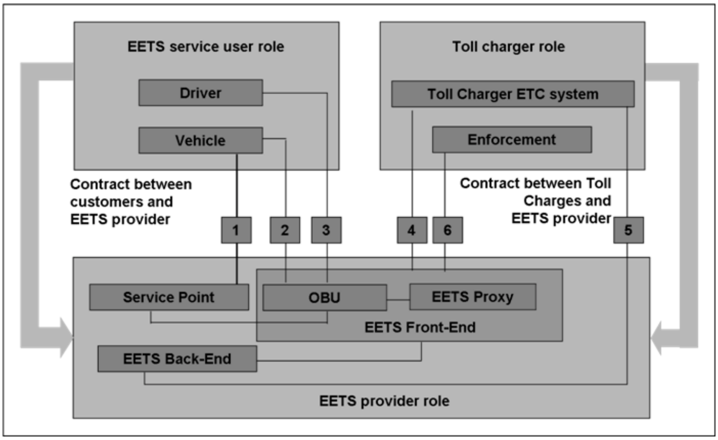


Fig. 1. RCI Project architecture [4]

RCI final report makes recommendations to [5] continue and finalize the standardization of the interfaces (CEN) and the work on the contractual aspects, define the technical EETS architecture and the interfaces, which are necessary for interoperability as elements in the EETS definition, determine the responsibility of the EETS Provider for the EETS Front-End (including the OBU) which must be stated very clearly in the EETS architecture, work with all stakeholders on a clear European roadmap of how progress will be made in the three years after the decision is finalized. This roadmap should make clear how the private sector can take its responsibility in the context of Member State action, European coordination and EC involvement.

The following interfaces should come standards:

- Interface 4 (Charging data according to ISO 12855, based on data Definition of ISO 17575-1, DSRC Charging – EN 15509; Localization Support – CEN TC278 WG1, SG6, PT 22/23).
- Interface 5 (Payment Data – ISO 12855, Security, Blacklists, Enforcement – ISO 12855, Toll Context Data – ISO 12855, based on data Definition of ISO 17575-3).
- Interface 6 (Enforcement for DSRC systems – EN 15509, Enforcement for GPS/GSM systems – ISO 12813).

---

[6] ASECAP is the European professional Association of Operators of Toll Road Infrastructures. It gathers and represents 17 Full Members (France, Italy, Spain, Portugal, Greece, Norway, Austria, Hungary, Croatia, Serbia, Belgium, the Netherlands, the United Kingdom, Poland, Denmark, Slovenia, and Ireland) and 4 Associate Members (Germany, Morocco, the Slovak Republic and the Czech Republic).

Interface 3 (HMI – Human Machine Interface) is required, but interfaces 1 and 2 are not applicable to CESARE role model and not critical for interoperability (however European consensus on high-level security requirements is needed).

4 The European Commission Decision of EETS

Based on the RCI program researches, Commission of the European Communities has implemented Decision on the definition of the European Electronic Toll Service (EETS) and its technical elements [1]. EETS sets out the necessary technical specifications and requirements for that purpose, as well as the contractual rules relating to EETS provision. Decision lays down obligations on EETS Providers, Toll Chargers and EETS Users. EETS domain means a toll domain falling under the scope of the 2004/52/EC Directive.

EETS Provider means a legal entity fulfilling the requirements and registered in a Member State where it is established, which grants access to EETS to an EETS User.

Toll Charger means a public or private organization which levies tolls for the circulation of vehicles in an EETS domain.

EETS User means a (natural or legal) person who subscribes a contract with an EETS Provider in order to have access to EETS.

On-board equipment means the complete set of hardware and software components required for providing EETS which is installed on board a vehicle in order to collect, store, process and remotely receive/transmit data.

Interoperability constituents means any elementary component, group of components, subassembly or complete assembly of equipment incorporated or intended to be incorporated into EETS upon which the interoperability of the service depends directly or indirectly, including both tangible objects and intangible objects, such as software.

As a minimum, the following standardized back office interfaces must be implemented by all EETS Providers. Toll Chargers must implement each interface, but can choose only to support either the GNSS or DSRC charging process:

- Exchange of toll declaration data between EETS Providers and Toll Chargers, specifically: submission and validation of claims for toll payment based on DSRC charging transactions, submission and validation of GNSS toll declarations;
- Invoicing / settlement;
- Exchange of information to support exception handling: in the DSRC charging process, in the GNSS charging process;
- Exchange of EETS blacklists and trust objects;
- Sending of Toll Context Data[7] from Toll Chargers to EETS Providers.

---

[7] Toll Context Data means the information defined by the responsible Toll Charger necessary to establish the toll due for circulating a vehicle on a particular toll domain and conclude the toll transaction.

5 National Automatic Toll Collect System Pilot Project

The Motor Transport Institute has created the structure of The National Automatic Toll Collection System for Poland (NATCS) in cooperation with FELA Management AG and AUTOGUARD SA. System consists of The National Automatic Toll Collection Centre (NATCC), on-board units (OBU) and control subsystem (Fig. 2).

The National Automatic Toll Collection Center (NATCC) based on telematics system approved all functions performed for the National Automatic Toll Collection System (NATCS). The NATCC elements are as follows:

1. Redundancy servers.
2. Applications and system software.
3. Data bases.
4. Interfaces: (between NATCC and OBU, between NATCC and control gates, between NATCC and external systems), User interface – www Internet service, Call Center, SMS gate, automatic telephone service.
5. Data transmission nets (WAN, LAN).

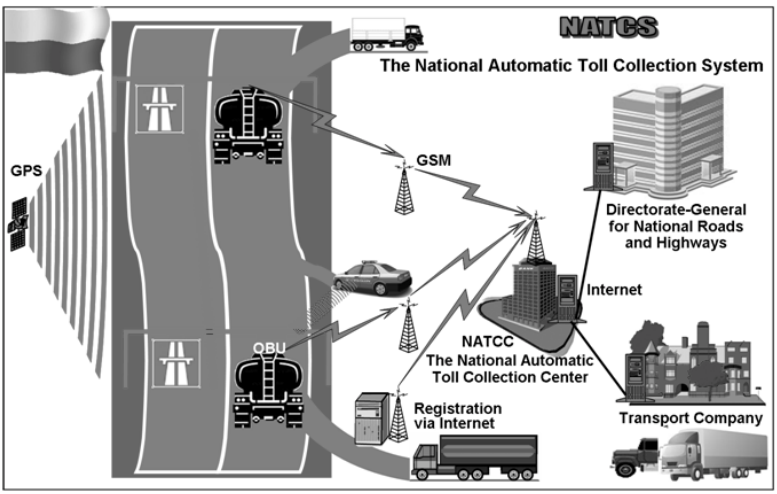


Fig. 2. The structure of the National Automatic Toll Collection System (NATCS)

OBU (Fig. 3) will be installed in vehicles windscreens and realized the following functions:

1. Vehicles data storage.
2. Digital map nodes and points storage.
3. Toll charges calculation based on introduced data (admissible mass, no of axles, emission class, distance, tariff model).
4. Analyzing the data coming from the module and sensors (GPS, GSM, and DSRC).

5. Optical and sound signalization of OBU working parameters (for instance distance, fees).
6. Safety data transmission to system and communication with control gates (stationary and mobile).
7. Remote data actualization and parameters exchange.
8. Data security based on cryptographic module.
9. Additional DSRC module to spread services and interoperability.

The system is based on an innovative combination of mobile telecommunications technology (GSM) and GPS, the satellite-based Global Positioning System. The main element of the automatic log-on system is the On-Board Unit (OBU). With the aid of GPS satellite signals and other positioning sensors, the OBU automatically determines how many kilometers have already been driven on the toll route, calculates the toll based on the vehicle and toll rate information that has been entered, and transmits this information to the NATCS computer centre for further processing.

Software will be supported with electronic road maps and data of users registered as well as data charges of highways and expressways.

Charge counting will start after highway entrance gate and finished after highway exit gate. Data on vehicle position will be additionally approved by GPS system and delivered to NATCC by GSM net. The toll amount is based on the truck's emission category and number of axles, as well as on the length of the toll route.

The control system distinguishes among automatic enforcement through control gates, enforcement by mobile teams and patrol teams. This combination guarantees comprehensive, continuous enforcement of the requirement to pay toll and allows the control system to be constantly adjusted to meet prevailing circumstances.

Automatic control subsystem consists of permanently installed enforcement control gates are used to ensure toll requirements are met without interrupting traffic flow (Fig. 3).

Fig. 3. Tripon EU OBU

6 Conclusions

According to European Commission the electronic toll collection systems in the European Union member states are not interoperable, so EC has taken important step to implement the 2004/52/EC Directive on the interoperability of electronic road toll systems in the Community.

Another important step is the decision on EETS definition and specifications. Decision to be the most important improvement for drivers since the abolition of border controls, stating that "the European Electronic Toll Service will enable road users to easily pay tolls throughout the whole European Union thanks to one subscription contract with one service provider and one single on-board unit.

The Commission Decision lays down the rights and obligations of toll chargers, service providers, and users. Users will be able to subscribe to the service provider of their choice. Toll chargers will communicate the tolls due to the service providers, who will eventually invoice the users. Tolls paid via EETS may not exceed the corresponding national or local tolls.

EETS will be available within three years for all road vehicles above 3.5 tones or allowed to carry more than nine passengers, including the driver. It will be available for all other vehicles within five years.

With regard to future expansion and development, the satellite-based toll collection system will be a better solution, especially with regard to flexibility when it comes to extending toll collection to every road category, every category of vehicle and, what's more, in terms of cost efficiency in implementation and operation.

The NATCS pilot project comes requirements of the 2004/52/EC Directive and European Commission decision of 6 October 2009. OBU is equipped with GPS, GSM and DSRC (5,8 GHz, IR) module to be interoperable with different type of EETS in member states. Equipment is designed in such a manner that its interoperability constituents utilize open standards and OBU is equipped with HMI, which indicates to the user that the OBU is functioning properly, and an interface for declaring variable toll parameters as well as for indicating the settings of those parameters.

Acknowledgements. This work has been supported by research project – NATCS – N R10 0001 04.

References

1. Černý, V.: Presentation on International Congres ITS, Prague (March 29 – April 1, 2009)
2. Commission Decision of 6 October 2009 on the definition of the European Electronic Toll Service and its technical elements. Official Journal of the European Union L 268/11 (October 13, 2009)
3. Directive 2004/52/EC of the European Parliament and of the Council of 29 April 2004 on the interoperability of electronic road toll systems in the Community. OJ of the EU, L 166/132 (April 3, 2004)
4. Schwarz-Herda, F.: Interurban road pricing. The Austrian Experience. Presentation to IMPRINT-NET group, Brussels (2006)
5. Springer, J.: Road Charging Interoperability. RCI project consortium, the European Commission DG TREN (November 2008)
6. The Act from 7 of November 2008 on changing act of public roads and some other acts. Official Journal of Polish Republic 2008, No 218, position 1391.Weiss, Andreas, Director Business Line Toll. EFKON AG. Presentation in General Department of National Roads and Motorways (May 8, 2009)

The Role of Websites in Promoting Urban Public Transport

Grzegorz Dydkowski[1,2] and Robert Tomanek[2]

[1] Komunikacyjny Związek Komunalny GOP, Barbary 21a, Katowice
[2] University of Economics in Katowice, 1 Maja 50, Katowice, Poland
tomanek@ae.katowice.pl, dydkowski@onet.pl

Abstract. Using Internet for the purpose of promoting mass transport has already become a fact. That fact results first of all from the continuous and widespread access to websites, the global reach of the Internet, as well as low unit costs of promotion via Internet. That is why the organizers and operators (services providers) of urban transport develop and modernize their web services – which as a result become better and better – which include not only the information indispensable during the journey (about the line routes and tariffs), but also the widely understood promotion of collective transport services.

Keywords: Public transport, promotion, websites of public transport.

1 Introduction

Promotion of collective urban transport is carried out by means of various instruments. A special place among them belongs to the websites of organizers and operators (services providers) of urban transport. A website serves as a source of information about transport, it is also a channel for distribution of services, as well as has a particularly vital role in the system of public relations [5].

Organizers and operators (services providers) of urban transport quickly appreciated the role of Internet in influencing transport behavior patterns. That is why a systematic development has been noted of websites devoted to urban transport. Internet services of the organizers are subject to appraisal – it can be noted that among the ones that were assessed as better has been the internet service of the Municipal Transport Union of the Upper Silesian Industrial Region (KZK GOP) in Katowice, Poland.

2 Directions and Examples of Making Use of the Website in Promotion of Public Transport

Promotion is a tool for communicating with consumers. Presently, when service providers take into consideration in their strategies not only the direct (final) buyers, but also the widely understood environment and stakeholders functioning in it, promotion should be understood as a process of communicating with the surroundings.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 133–141, 2010.

Among the elements of promotion one can include:

- advertising, which is a ”paid form of impersonal marketing communication concerning the organization, product, service, or idea, by an identified sender” [7],
- direct personal sale, defined as ”bilateral information flow between the purchaser and seller, having the purpose of influencing individual or group purchase decision” [7] – including, in increasing way, the electronic channels,
- public relations – ”all activities having the purpose of promoting and/or protecting the image of an enterprise or product” [3],
- promotion of sales, that is ”short-term display of the value offered, as invitation to purchase given goods or services” [7].

Practically speaking, in case of each of the elements mentioned Internet, especially websites, plays an ever increasing role in collective urban transport. The only exception there is advertising – as in Poland one can relatively rarely encounter advertising of urban transport services by means of Internet. On the other hand, in the case of sales promotion and public relations websites seem virtually irreplaceable. Also a direct personal sale, thanks to selling the so-called virtual tickets (tickets used by means of cellular phones) becomes a promising area of promotion activities in urban public transport.

Promotion activity encountered at websites of organizers and operators (services providers) of collective urban transport takes the form of integrated projects, oftentimes implemented by several different entities (organizers of transport, or organizers and units from outside the sphere of collective transport – e.g. police).

Among the examples, one can list here:

- image-generation campaigns for urban transport (and entities that provide such services), which have the purpose of changing transport-related behaviour and improving the reputation of the provider of transport services – such activities may be included in public relations activities, but also in sales promotion (in the light of the definition provided above), as sometimes they also bear the signs of sales promotion,
- passenger information systems, which may be classified within the sphere of direct sales,
- sales of tickets via Internet (also direct sales).

Image-generation campaigns become a common way of forming the reputation of urban transport, and are used ever more widely in developing the division of transport tasks. Sometimes they have the form of nation-wide campaigns, but usually are of local character. The most popular in Poland was the social campaign ”Swap the car for a bus” („Zamień wóz na bus”) initiated in 2007 by the Urban Transport Chamber of Commerce (Izba Gospodarcza Komunikacji Miejskiej - IGKM) from Warszawa. Participating in the campaign were transport organizers and operators (providers) from 15 towns. The climax of the campaign was participation in the European ”day without car”. Within the framework of the campaign, press conferences were organized – the aim of all activities was to stress the advantages and role of public transport in the development of sustainable transport in cities [16]. It is a striking thing that a campaign which involved huge areas and provided interesting promotion solutions is not continued in such a form. That seems to indicate a poor integration of the Polish mass transport community. Although the chamber IGKM exists yet, despite many years of its activities, that organization appears unable to follow such examples as that presented by VDV (Verband Deutscher Verkehrsunternehmen), active in Germany [17]. VDV performs actual active marketing functions concerning image building (also in reference to cargo transport). One of the examples of the activities of that association is the campaign, initiated in 2007, under the motto ”Do something for climate – develop your rolling stock” [15]. VDV angages in many more such campaigns, whereas in Poland they have local character – it seems that this calls for changes.

The Municipal Transport Union of the Upper Silesian Industrial Region (KZK GOP) has interesting experience in social (image-building) campaigns, which union not only took part in the campaign ”Swap the car for a bus” („Zamień wóz na bus”) (with the aim of changing transport-related behaviour), but also engages in its own social campaigns [6]. Those initiatives have the following aims:

– changing transport-related behaviour,
– improvement of the image of urban transport organizer (KZK GOP),
– enhancing safety (not only in the means of collective transport).

One of the most original initiatives of KZK GOP concerning improvement of the image is the campaign called ”Let us meet on the bus” and ”Let us meet on the tram”, the purpose of which is pointing out the importance and difficulties of the work of ticket inspectors. That campaign has been continued since 2008, and consists of displaying posters at stops, promoting the work of ticket inspectors of KZK GOP [12]. A similar aim exists in the case of campaign ”Show your ticket, do not be ashamed”, which promotes a new solution concerning organization, consisting of showing tickets to the driver upon entering the vehicles on selected lines [11]. An interesting area of image-related activities of KZK GOP, popularized via the website of that transport organizer comprises the efforts to improve safety (not only in public transport, but also in urban space environment). KZK GOP has been involved in activities in that respect, among others in co-operation with the police. The main campaign, launched in 2007, is that called ”It is safer together with KZK GOP and Sznupek” – the main character or in fact protagonist of the campaign is Sznupek (a plush mascot), which teaches children the principles of safety and of using urban transport. That campaign puts together the elements of safety and development of a positive image of collective transport and the ability of using transport services [13]. Image-oriented initiatives of KZK GOP are presented at the website of that transport organizer, it is also a channel for contact of service recipients with KZK GOP. That website serves the purpose of publicizing the campaigns, and communicating with potential recipients.

The website is also a basic tool for providing transport-related information. In particular, the following information is under consideration:

– information on tariffs (including promotions),
– information on time tables,
– recent/ current information about changes in the transport offer.

As concerns tariffs, of importance for the users of services are particularly the information concerning fares, eligibility of reduced fares and using transport free of charge, the procedures of appeal from the penalties imposed, as well as zones in the tariff structure.

Among the systems of providing information on transport, worth distinguishing are the systems of presenting time-tables and journey planning. Internet allows not only to provide exceptionally simple and cheap access to information about transport. At present, organizers and operators (service providers) very frequently provide time-tables on their websites, in the form of time-tables for specific stops and for specific lines/ routes. Even complex systems of routes and lines, characteristic for big conurbations, are presented in a digestible, convenient form. This has already become a standard for the websites of collective urban transport. Less common are the so-called journey planners, allowing planning a journey using various lines/ routes and pedestrian access. The flexible solution called Google Transit [4] is used even more often, which may be linked with websites of urban transport organizers. As of 18.04.2010 there are five cities in Poland, which have been using such a solution: Białystok, Olsztyn, Szczecin, Warszawa, and Zielona Góra, compared with 446 cities worldwide [9].

Sales of tickets via Internet are still rare in Poland, although those systems are currently intensely developed. Solutions are introduced which provide the possibility of paying via Internet for the so-called mobile tickets (used in cellular phones) or – in case of charging by means of electronic cards – collecting the toll, while the content provided on the card is updated during the next use of collective transport. That is in line with the trends present in other countries, where that channel of distribution is developed (e.g. Vienna [10], Helsinki [14]). In the conditions of e-economy, which has been becoming the reality of the 21st century, the introduction of sales of services via Internet is a particularly important issue [1] and [2]. One can expect that the number of people will be increasing, for whom the use of public transport will depend upon the possibility of easy, electronic purchase of the ticket. The barrier for buying tickets via electronic means in Poland is the necessity of registering beforehand and pre-paying sufficient funds, or agreeing for debiting one's bank account, as the purchased tickets are not settled with payments for telecommunication services. Also the payments for the fare and for data transmission that took place during the purchase are settled separately.

3 Assessment of Functionality of Websites of Urban Transport Organizers in Poland

The basic measure of functionality of a website is the traffic generated there, measured by the number of visits, taking into consideration more detailed data – including which parts of the website are visited most frequently, and how the number of visits changes with time. For example, the home page of KZK GOP website, between May 1, 2009 and April 30, 2010, was visited nearly 12.5 million times, Fig. 1.

Many more visits – nearly 141.5 million – were registered for pages related to the time-table, which were the most often viewed pages at the KZK GOP website. Users usually made use of the web pages of the Union KZK GOP several times. Depending upon the content, some 71% to 78% of users returned to KZK GOP website, while other visits were one-off events.

It results from the above that the home page of KZK GOP was opened over 34 thousand times a day, on the average, whereas the pages with time tables had nearly 388 thousand visits a day. The number of visits changes in specific months. The peak activity of internet users visiting KZK GOP web pages was recorded in October 2009, whereas the opposite was true for holiday months. Also the time spent by Internet users at KZK GOP web pages differs depending on the page content. The time of displaying the home page varies between 47 seconds and 1 minute and 52 seconds, whereas the time spent on the pages related to time tables amounts to between 3 minutes and 3 minutes and 18 seconds.

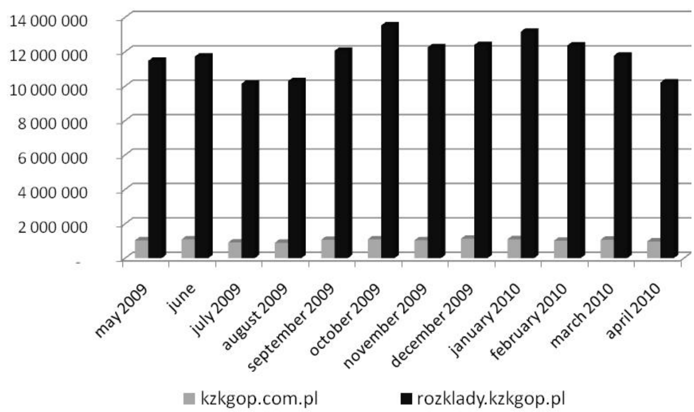


Fig. 1. Number of visits to the home page of KZK GOP and pages related to time-table, in the period May 1, 2009 - April 30, 2010. Source: Own study on the basis of KZK GOP data.

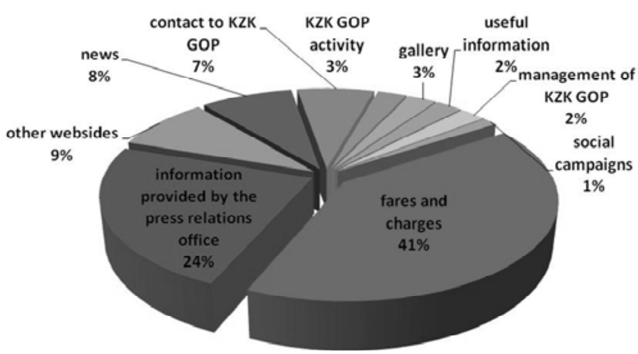


Fig. 2. Distribution of visits to KZK GOP web pages, apart from the home page of KZK GOP and pages related to time-tables, over the period: 1.1.2009 -30.4. 2010. Source: Own study on the basis of KZK GOP data.

In other words, the parts of the KZK GOP website, besides those pages which are devoted to time tables, which are most often used by internet users, are the pages referring to fares and charges, information provided by the press relations office, news, and contact to employees of the Transport Union. Fig. 2.

The number of internet users visiting the website of KZK GOP varies in specific periods. It can be gathered from Fig. 3 and Fig. 4 that the number of the home page users of the Transport Union is on the decline.

However, there is a steady increase in users that visit pages related to time tables. It is then concluded that the people who want to obtain the necessary information go directly to the page of their interest, without visiting the KZK GOP home page. Those are probably regular users of the KZK GOP website.

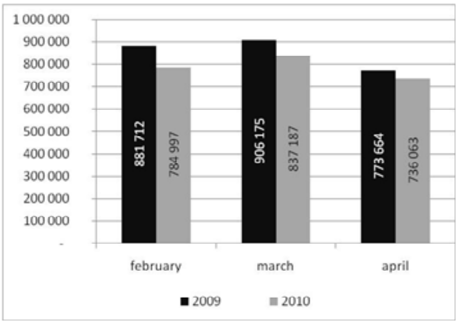


Fig. 3. Dynamics of visits paid to the KZK GOP home page between February and March, 2009, and 2010. Source: Own study on the basis of KZK GOP data.

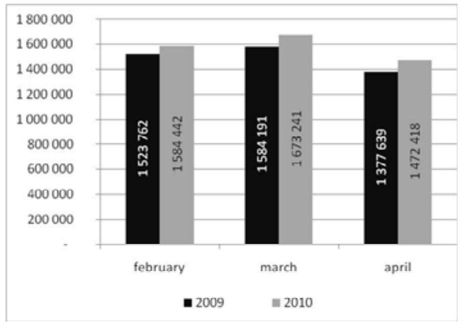


Fig. 4. Dynamics of visits paid to the KZK GOP pages related to time tables, between February and March, 2009, and 2010. Source: Own study on the basis of KZK GOP data.

The efficiency of use of the website in promotion depends upon numerous factors, in particular upon the functionality and aesthetics of the website itself. In Poland, cases of websites of urban transport organizers and operators (service providers) benchmarking are rare, one of the few made is the audit performed in 2009 on ten biggest service providers for urban transport in Poland – the study was performed by EDISONDA [8]. The following elements of internet services were examined: home page (weight 1), time tables – availability (weight 3), time tables - readability (weight 3), connection search (weight 3), maps and schemes (weight 1), fares (ticket prices) (weight 1), changes in time tables (weight 2), navigation in the service (weight 2).

The audit was based on a point scale, where specific categories could obtain a score between 0 and 5 points. Specific scores were weighted using weights from 1 to 3. The weight values have been provided above. Of course, those weights may raise concerns of doubts – yet that is the right of the authors of the audit, which would be difficult to dispute.

As regards the home page, attention was paid to such elements as aesthetics and readability of the page. In particular, the factors that negatively influence the appraisal of the home page include a large number of banners, as well as a complicated menu. Of importance is also a suitable division of the page, colours, and structure of the navigation menu. The page/ website should not have a complex structure – access to the lowest level of information should not require more than three clicks. The examined websites of service providers in urban transport appear to show some differentiation, and received marks from 1.5 points (Łódź) to 4 points (Warszawa, Wrocław, Poznań). Those websites change year by year, yet it is noticeable that still a lot of work is required to make them modern, so that they meet the needs concerning promotion of urban transport services. The following solutions may be recommended for the home page:

- restricting the amount of information to a minimum (the page should not distract attention) and providing such organization of the page that it is not necessary to use scroll-bars (e.g. thanks to the application of a "building blocks" structure),
- organization of the menu having in mind the division into information required for service users, as well as information about the service provider (addressed mainly to media and suppliers of the operator (service provider)/organizer),
- direct access to information about time tables (network).

An important element of the urban transport websites is the information about time tables. Usually they have the form a list of lines (routes). The availability of time tables varied (being the best in Gdańsk – 5 points, and the poorest in Wrocław and Szczecin – 2.5 points each). As regards the readability of time tables, the highest score was that of KZK GOP (5 points), and the lowest that of Bydgoszcz (1.5 points), where the information was "blurred" by providing the information in 3 languages simultaneously. It is worth stressing here that in case of information provided by KZK GOP, when constructing the website an additional difficulty had to be overcome, namely the complexity of the transport network, comprising not only the 25 municipalities belonging to the union, but also several adjacent municipalities. Among the recommendations concerning the provision of information about the line, the following can be listed:

- explicit navigation to the page with time tables,
- access to information about time tables by searching not only the numbers of lines, but also stops,
- publication of additional information about specific services,
- time tables should be easily printable (also in shortened versions).

Search engines for connections constitute a particularly valuable functionality of websites. They are indispensable in conurbations. Independent of the Google Transit function presented before, it is necessary for the biggest cities to offer such a functionality in their internet services. In the case of the studied ten biggest cities in Poland, it was only Lublin that did not offer that possibility. The other cities had search engines – the highest score for search engine was achieved by KZK GOP (5 points), the lowest by Wrocław (3 points), still those solutions were generally relatively highly appraised. Similarly as in the case of information about time tables, also there it is necessary to provide easy access to the search engine (the more so, that it is often used by guests to the area, who did not have the opportunity earlier to use the services of a given transport service provider) as well as easy legibility and explicitness of information generated by the service.

Internet services in urban transport must also enable easy access to information about tariffs, as well as changes in time tables (traffic limitations). In particular, one should note here the necessity of limiting the use of specialist vocabulary. Appraisal of specific services from the point of view of availability of information about fares is high, and quite similar (the highest scores in Warszawa and Poznań – 4 points each), the lowest (2.5 points) in Gdańsk. On the other hand, more problematic is the information about changes in time tables – due to the many changes (caused mainly by numerous repairs of road infrastructure) it appears difficult to provide the necessary information in easily readable fashion. That is why scores are lower and vary from 3.5 points (Warszawa and Łódź) to 1.5 points (Wrocław). It seems desirable that is case of changes to time tables, information about such changes should be published already in the time table plan, and that changes are marked by different colours, while the home page should signal changes and limitations that occur.

The navigation in such services was not appraised too high (with the exception of Warszawa, scoring 4.5 points). It would be advisable to provide the possibility of returning to the home page from every level of viewing, and to provide access to an abridged map of the site as specific pages. The results of appraisal of websites according to the criteria discussed are provided in Table 1. As it can be seen, the highest scores were that of KZK GOP, slightly ahead of ZTM Warszawa. One can see here the differentiation between specific service providers, as regards the appraisal of the internet services offered.

Table 1. Appraisal results for websites of urban transport in Poland

Category / city (region)	**[weight]**	**Warszawa**	**Łódź**	**Kraków**	**Wrocław**	**Poznań**	**Gdańsk**	**Szczecin**	**Bydgoszcz**	**Lublin**	**GOP**
home page	1	4	1,5	3,5	4	4	2,5	3	2,5	3	3,5
time tables-availability	3	10,5	12	12	7,5	13,5	15	7,5	10,5	10,5	13,5
time tables- readability	3	10,5	10,5	10,5	7,5	10,5	13,5	7,5	4,5	12	15
connection search	3	12	12	12	9	12	12	9	10,5	0	15
maps and schemes	1	4,5	4,5	5	0	4	4	4	2,5	1	0
fares (ticket prices)	1	4,5	3,5	3	4	4,5	2,5	4	4	3	3,5
changes in time tables	2	7	7	6	3	4	4	5	5	5	5
navigation in the service	2	9	6	7	4	7	7	5	6	5	7
Weighted average		62	57	59	39	59,5	60,5	45	45,5	39,5	62,5

Source: [8].

Analyzing the results of the ranking discussed, one can note that websites are still not fully utilized as a tool for promoting collective urban transport. In particular, numerous changes are needed concerning the easiness of navigation and making use of time-table information. Corrections are also necessary as regards the offered search engines for finding connections. It is evident that websites require continuous modifications – they are by no means established once and for good, and the provision of information, as well as compactness of structure, along with mechanisms of providing services, call for incessant modifications.

4 Conclusion

Use of Internet in the promotion of collective transport became necessary. At the same time, it poses a chance for efficient promotion of collective transport, in all dimensions of promotion (not only advertising, but also personal direct sale, public relations and sales promotion). Organizers and operators of urban transport in Poland are aware of that fact, and take ever more efforts to develop further and to modernize their Internet services. One can expect this tendency to be durable. It would be advisable to use websites in a wider manner for promoting the changes in transport-related behaviour – in particular through wide-spreading their image-building campaigns. It is worth pointing out the necessity of integrating the efforts of organizers and operators (providers) of collective urban transport. A platform for such integration should be a nationwide organization (following the example of the German VDV) with membership of transport organizers and operators, as well as non profit entities interested in the development of collective transport. Changes in the functioning of the Polish chamber of urban transport (IGKM) appear to be necessary; alternatively the service providers may establish a new entity of such kind.

References

1. 7th UITP Conference on Automatic Fare Collection Towards a Single Market for Fare Collection. Conference Proceedings (material in electronic form - CD), Bologna (2004)
2. IT Solutions for Public Transport. Conference Proceedings, (material in electronic form CD), Karlsruhe (2008)
3. Kotler, P.: Marketing. Analiza, planowanie, wdrażanie i kontrola. Gebethner i Ska, Warszawa (1994)
4. Krzaczkowski, Ł.: Google transit - nowe podejście do informacji pasażerskiej. "Biuletyn Komunikacji Miejskiej" 94 (2007)
5. Marketing Best Practice. UITP Commission on Marketing & Product Development. UITP (2003)
6. Promocja, kampanie, konkursy społeczne KZK GOP, Katowice (2009) (unpublished materials in electronic form)
7. Przybyłowski, K., Hartley, S., Kerin, R., Rudelius, W.: Marketing. Dom Wydawniczy ABC (1988)
8. Turaj H.: Serwisy internetowe komunikacji miejskiej. "Komunikacja Publiczna" No. 1 (2010)
9. `http://mapy.google.pl/intl/pl/landing/transit/text.html#eu` (April 18, 2010)
10. `http://shop.wienerlinien.at/` (April 18, 2010)
11. `http://www.kzkgop.com.pl/informacje/archiwum/2008/66/index.php` (April 18, 2010)
12. `http://www.kzkgop.com.pl/kampanie/kanarek/` (April 18, 2010)
13. `http://www.kzkgop.com.pl/kampanie/sznupek/` (April 18, 2010)
14. `http://www.plusdial.com` (April 18, 2010)
15. `http://www.vdv.de/oepnv_themen/umwelt.html?pe_id=165` (April 18, 2010)
16. Witryna internetowa, `http://www.zamienwoznabus.igkm.pl//index.php?option` (April 15, 2010)
17. Wolański, M.: Odczarowywanie komunikacji- inspiracje i możliwości, Warszawa (2008) (unpublished materials in electronic form)
18. Mikulski, J.: Wizja rozwoju inteligentnych systemów transportowych w Polsce. Magazyn Autostrady 10, 34–40 (2009)

Assisted-GNSS, Why, Where and for Whom?

Jacek Januszewski

Gdynia Maritime University, al. Jana Pawla II 3
81-345 Gdynia, Poland
jacekjot@am.gdynia.pl

Abstract. As in restricted areas, in urban areas in particular, the accuracy of the user's position obtained with satellite navigation systems (SNS), e.g. GPS, may be considerably poorer than in open areas, and in some cases cannot be obtained, other solutions for position fixing must be used. An Assisted-GPS (A–GPS) is now one of these methods – a new technology that uses an assistance server to cut down the time needed to determine a location using GPS (Time To First Fix – TTFF); A–GPS can be used in all SNS today and in the future (GNSS). The principle of this method, and where, when and for whom it can be used is described in this paper.

Keywords: Global Satellite Navigation System, Assisted–GNSS, satellite visibility, urban canyon.

1 GNSS Limitations

In May 2010 a terrestrial position fix can be obtained with Satellite Navigation Systems (SNS) such as the GPS (the only fully operational) and GLONASS, in the future also with Galileo and Compass. All these systems are known as the Global Satellite Navigation System (GNSS).

SNS receivers obtain signals from satellites and use the data to calculate a 3D position (latitude, longitude, altitude and time) or 2D position (latitude, longitude and time). That is why to operate properly, a conventional SNS receiver needs a clear view of the skies from at least four or three satellites before it can calculate its position. This requirement excludes fully the operation in unfriendly conditions, such as in buildings, in urban "canyons", under heavy tree cover or other shadowed environments or even indoors. With an unassisted SNS the Time to First Fix (TTFF) including the process of locating the satellites, receiving the data and achieving a position fix, can take several minutes. The expression TTFF generally refers to the time needed by the receiver to perform the first position fix, starting from the moment it is switched on. This delay can be problematic for many SNS applications, for emergency in particular. Additionally the accuracy of the user's position can be compromised and in same cases it can be impossible to achieve a position fix.

The percentage of satellites visible by the user above given masking elevation angle H decreases with this angle for all SNSs and for all user's latitudes. If the angle H is equal 0° we can say that 100% of satellites visible can be used by the user, if this angle is equal 20° the percentage of these satellites decreases to 60%, if H = 40° this

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 142–151, 2010.

Table 1. Mean number of satellites l_{ms} visible above H_{min} and obstacles to the observer situated in the middle of the street (width L = 70 m, height B = 15 m) for different angles H_{min}, for different angles between the North and street axis (α), for Galileo and GPS systems at different observer's latitudes φ; l_m – mean number of satellites visible above H_{min} in an open area

φ [°]	H_{min} [°]	System	l_m	Angle α [°] 0		90	
				l_{ms}	l_{ms}/l_m [%]	l_{ms}	l_{ms}/l_m [%]
0 – 10	0	GAL	11.05	7.77	70.3	7.34	66.4
		GPS	10.74	7.54	70.2	7.04	65.5
	5	GAL	10.02	7.77	77.5	7.23	72.2
		GPS	9.75	7.54	77.3	7.00	71.8
	10	GAL	8.96	7.74	86.4	7.12	79.5
		GPS	8.69	7.51	86.4	6.85	78.8
	15	GAL	7.84	7.35	93.8	6.83	87.1
		GPS	7.59	7.09	93.4	6.57	86.6
50– 60	0	GAL	10.85	7.62	70.2	7.23	66.6
		GPS	10.40	7.00	67.3	6.89	66.3
	5	GAL	9.67	7.51	77.7	7.17	74.1
		GPS	9.15	6.86	75.0	6.84	74.8
	10	GAL	8.35	7.12	85.3	6.99	83.7
		GPS	7.90	6.56	83.0	6.66	84.3
	15	GAL	7.14	6.62	92.7	6.66	93.3
		GPS	6.80	6.21	91.3	6.36	93.5
80– 90	0	GAL	11.28	8.53	75.6	8.50	75.4
		GPS	10.90	8.16	74.9	8.12	74.5
	5	GAL	10.38	8.48	81.7	8.44	81.3
		GPS	9.35	8.10	86.6	8.03	85.9
	10	GAL	9.46	8.28	87.5	8.28	87.5
		GPS	9.07	7.91	87.2	7.87	86.8
	15	GAL	8.53	7.95	93.2	7.96	93.3
		GPS	8.16	7.57	92.8	7.57	92.8

percentage is about 30% only. That is why the possibility of the SNS use in restricted areas is very limited [1], [2], [3], [4].

In restricted areas, in urban areas in particular, the accuracy of the observer's position obtained with the SNSs depends on the number of satellites (ls) visible above masking elevation angle (H_{min}), the geometry of systems – the Geometric Dilution of Precision (GDOP) coefficient, the dimensions (the width of the street and the building height) & location of the obstacles (the angle between the North & street's axis and user's latitude).

Table 2. Mean number of satellites visible above H_{min} = 5° and the obstacles by the observer situated in the middle of the street for different widths L and different heights B in the zone 50–60° for Galileo and GPS systems, the street axis in the North–South direction.

B [m]	System	L [m]						
		10	20	30	40	50	60	70
5	GAL	4.04	6.96	8.18	8.73	9.09	9.31	9.46
	GPS	3.71	6.27	7.63	8.25	8.59	8.79	8.93
10	GAL	2.18	4.04	5.73	6.96	7.71	8.18	8.50
	GPS	–	3.71	5.11	6.27	7.08	7.63	8.00
15	GAL	–	–	4.04	5.16	6.20	6.96	7.51
	GPS	–	–	3.71	4.65	5.53	6.27	6.86
20	GAL	–	–	–	4.04	4.89	5.73	6.43
	GPS	–	–	–	3.71	4.43	5.11	5.73
25	GAL	–	–	–	–	4.04	4.72	5.41
	GPS	–	–	–	–	3.71	4.29	4.85

As an example the mean numbers of satellites (l_{ms}) visible above H_{min} and the obstacles blocking the observer situated in the middle of the street for different angles between the North and street axis (angle α) for GPS and Galileo systems for three selected zones of latitude (two extreme 0–10° and 80–90° and zone 50–60°, latitude interval of Poland) are demonstrated in the Table 1. The calculations were made for four angles H_{min} (0°, 5°, 10° and 15°) for width L = 70 m, height of buildings B = 15 m, for two angles α (0° and 90°). Additionally the ratio l_{ms}/l_m (l_m – the mean number of satellites visible above H_{min} in an open area) in per cent is also shown. Let us recapitulate:

- the number l_{ms} depends on the system; in this case for the Galileo system is always greater than for the GPS system,
- the number l_{ms} depends on the observer's latitude for each angle H_{min} for each angle α for both systems. This number has a maximum in zone 80–90° and a minimum in zone 50–60°,
- the number l_{ms} decreases and the ratio l_{ms}/l_m increases with angle H_{min} in each zone for each angle α for both systems,
- in zone 80–90° the number l_{ms} for different α is practically the same for both systems for each H_{min}, in other zones l_{ms} depends on angle α in each case.

The additional calculations were made for different width L and different height B in the zone 50–60° for street's axis in the direction North–South (α = 0°) and in the direction West–East (α = 90°). Mean number of satellites l_{ms} visible above H_{min} = 5° (a masking elevation angle used in most receivers) and the obstacles for the observer situated in the middle of the street for Galileo and GPS systems for α = 0° is presented in Table 3. We can say that:

- the number l_{ms} for both systems increases with width L and decreases with height B for both systems. For each width L there is a critical value when l_{ms} is less than 4 and the position fix in mode "3D" cannot be obtained,
- as the number l_{ms} for the Galileo is always greater than for the GPS, it means that for given values of L and B the Galileo "3D" position fix can be obtained, while the GPS cannot,
- the number l_{ms} for both systems depends on the angle α, however this dependence is greater for smaller values of L and B. It means that if the axis of the street runs in West–East direction the position fix can be obtained, while in the direction North–South cannot (i.e. for L = 25 m and B = 15 m).

These results and these two tables, in particular, show that the possibility of the position fix with all GNSS in urban areas depends on many elements, and very often is limited, sometimes even impossible [3], [4].

2 Why Assisted–GPS ?

The U.S. Federal Communications Commission (FCC) in 1996 ordered that cellular phone carriers had to route 911 calls to a Public Safety Answering Point (PSAP) with Automatic Location Identification (ALI). It was recognized that, at that time, there was no technology that could provide a detailed caller–location information. The solution as a handset with a GPS receiver incorporated seemed excluded, because [5]:

- after lauching a 911 emergency call it would take the receiver several minutes to establish a fix and being able to output the caller's location coordinates; it means that the TTFF is too long;
- a GPS receiver would drain too much power from the handset;
- the GPS system would have a very poor performance in restricted areas;
- incorporating a GPS receiver in the handset would be far too expensive and too bulky.

In many poor signal–to–noise environments, the signals from satellites can be acquired and tracked, but the navigation data message cannot be demodulated. Therefore, a stand–alone GNSS receiver may have to rely on out-of-date ephemeris, satellite clocks, and ionosphere calibration parameters, degrading the navigation solution, while a "cold start" navigation solution cannot be obtained at all [6], [7].

Standard stand-alone GPS receivers take one to two minutes to search for acquired satellites, but they cannot lock on them if the signals are attenuated, as it occurs in restricted areas, in indoor environments in particular. In fact, when performing a cold start, a standard GPS receiver must do a search both in frequency and in code space with only two correlators per channel, with which they can search sequentially the 1,023 possible code delay chips, in each adjacent frequency bin [8].

Finally, we can say that in the GPS receiver acquisition it is easier if four things are known: the frequency offset, accurate time, code delay and receiver position. The last element affects the observed code delay and the observed satellite Doppler. However, a conventional GPS receiver needs to acquire and track a signal before it can find any of this information. This conundrum leads to the fundamental idea of GPS assistance;

providing the GPS receiver with all the information you can, by some alternative means of communications [9].

3 What Is an Assisted GPS and for Whom ?

The technology which provides now a detailed caller–location information in the navigation data message is an Assisted GPS (A–GPS), also known as a network assistance, which uses a separate communication link, such as mobile phone system. The A–GPS uses a GPS reference network, formed by many A–GPS servers or location servers, connected to the cellular infrastructure and consisting of GPS receivers than can detect GPS signals continuously and monitor the satellite constellation in real time.

Because of the long integration times needed for acquisition and the difficulties in decoding the navigation message from a satellite, a successful weak signal positioning depends considerably on an outside assistance information. We can distinguish three main assistance ingredients to acquire the signals and to compute the user's position: the satellite orbits, time and receiver location needed consecutively to calculate satellite position, make pseudorange measurements and predict the satellite Doppler shift in acquisition. The A–GPS can be employed within a wireless network. This network can send the GPS receiver, for example, a prediction of the received signal including the Doppler shift, code delay, initial position estimate, and the navigation message [10].

Additionally A–GPS architectures increase the capability of stand-alone receivers to save the battery power, acquire and track more satellites, thereby improving the observation geometry, and increase the sensitivity against a conventional GPS architecture. These enhanced capabilities result from the knowledge of the satellite position and velocity, the initial receiver position, and the time supplied by the assistance server [11].

In regular GPS networks there are only GPS satellites and GPS receivers. In A–GPS networks, the receiver, of limited processing power and normally under less than ideal locations for position fixing, communicates with the assistance server that has a high processing power and access to a reference network. Since the A–GPS receiver and the Assistance Server share tasks, the process is quicker and more efficient than a regular GPS, albeit dependent on the cellular coverage [12].

The reference network has several tasks, such as providing an approximate position, satellite parameters, Doppler data, and others. The assistance information transmitted from the GPS reference network to the user achieving the following benefits [8], [13]:

- the reduction of time needed to determine the position, i.e. the TTFF from approximately 30 seconds to a few seconds;
- enhancing the sensitivity of a GPS receiver, allowing it to determine the position even in restricted areas;
- improving the position accuracy to the level obtained with a differential GPS.

A stand-alone GPS receiver must search for satellite signals and decode the satellite navigation messages before computing its position – tasks which require strong signals and additional processing time. It means that in this receiver the phase of acquisition of GPS signals is practically a search over the whole possible frequency and code-delay space.

The received GPS signals are shifted in frequency due to the relative receiver–satellite motion. This is so-called Doppler frequency shift. The receiver must find the frequency of each signal before it can lock onto it. The knowledge of the satellite position and velocity data and the initial receiver position reduces the number of frequency bins to be searched. This reduction is possible because the receiver can directly compute the Doppler frequency shift instead of searching over the whole possible frequency range. Additionally the Doppler shift on the PRN code of the GPS receiver must be taken into account. The satellite position and velocity data are computed from the orbit and clock data provided by the assistance server. The initial receiver position can come from cellular techniques or any other available source of information.

The reducing of the number of frequency bins which must be searched to acquire the signal reduces the TTFF. This shorter time results in a reduced power consumption because the system does not have to wait for the GPS receiver to decode the navigation data for each visible satellite. If the receiver had to decode the ephemeris from the broadcast message, it would take a minimum of 18 seconds after acquiring the signal, assuming that it did not drop or lose any data bits [11].

In an A–GPS the outside sources, such as an assistance server via a network, help a GPS receiver to perform the tasks required to make range measurements and calculate position solutions. The assistance server has the ability to access the information from the reference network and also has the computing power far beyond that of the GPS receiver [12]. The A–GPS server using Long Term Orbits (LTO) supplies the future ephemeris data to handsets that is valid for up to a week. This enables the benefits of A–GPS technology when temporarily out of mobile operator network range.

Two other advantages of A–GNSS are a greater sensitivity and customer satisfaction. Because the receiver has fewer frequency bins to search, it can dwell in each bin for longer periods of time. As this additional dwell time increases the sensitivity of the receiver, it can use signal strengths below the conventional thresholds to make the range measurements. In the recapitulation we can say that A–GNSS allows the use of satellite data which would have otherwise been unavailable. The second advantage is the customer satisfaction when using a location or emergency call. If the position fixing takes minutes, as it is common with warm starts in conventional GNSS receivers, the consumer might become frustrated while waiting and wonder whether there was anything wrong with the phone. With A–GNSS the user's position can be computed more quickly, within a few seconds [11].

The A–GNSS can be used if all these areas, where stand-alone GNSS receivers are not sufficient in severely obstructed signal conditions to provide means to obtain the location. The largest application of A–GNSS, today A–GPS, is in mobile phones, where the provider of the assistance data is usually the network operator supporting mobile phones from many different manufactures [9], e.g. BCM 4750 – single chip AGPS solution (Broadcom), UBX–G5010 – single chip version of high performance u-blox 5 positioning engine (U-blox).

4 Basic Methods of Assisted GPS

There are two basic methods (called also solutions or operation) of assisted GPS employed in cellular handsets:

- User Equipment (UE) – assisted or Mobile Station (MS) – assisted GPS,
- UE – based or MS – based GPS.

These two methods are quite different, but both require a complete or nearly complete GPS receiver to be integrated into the cellular handset. The defining difference is where the mobile station position is calculated. This difference also affects the assistance data. The technical description and performance of A–GPS positioning are presented in Table 3, the handset input and output data messages corresponding to both methods in Table 4.

Table 3. Assisted–GPS positioning, the global characteristics [14]

Technical description	Performance
• technology – radio frequency in the 1.575 GHz and in the mobile network range • localization – infrastructure or mobile, quasi-continuous • positioning – absolute • environments – indoors, outdoors	• accuracy – from 10 to 100 m • range – mobile network coverage

Table 4. Assisted GPS – exchanged information elements between the A–GPS server and the user equipment (UE) [8] and [15]

	MS–based	MS–assisted
From A–GPS to UE (handset inputs)	• visible satellite list • satellite-in-view signal Doppler and code phase • ephemeris • approximate position • GPS reference time • almanac • DGPS corrections • real time integrity (failed/failing satellite information) • sensitivity assist	• visible satellite list • satellite-in-view signal Doppler and code phase or, alternatively, approximate handset position and ephemeris • GPS reference time
From UE to A–GPS (handset outputs)	• location request • coarse position • calculated position • velocity, time	• location request • coarse position of the user equipment • time-stamped GPS pseudoranges
Position computed	in user equipment (UE)	in the network

Figure 1 illustrates the structure of both mentioned A–GPS methods.

The MS–assisted solution saves the computing power and memory at the MS, whereas the MS–based solution can also be used as a stand-alone GPS receiver. The MS–based solution has relatively short uplink information elements (IE), whereas downlink assistance IEs are relatively long; the opposite is true for MS–assisted. Depending on the type of the location request, the position can be returned from the network to the MS, or vice versa [8], [9], [15].

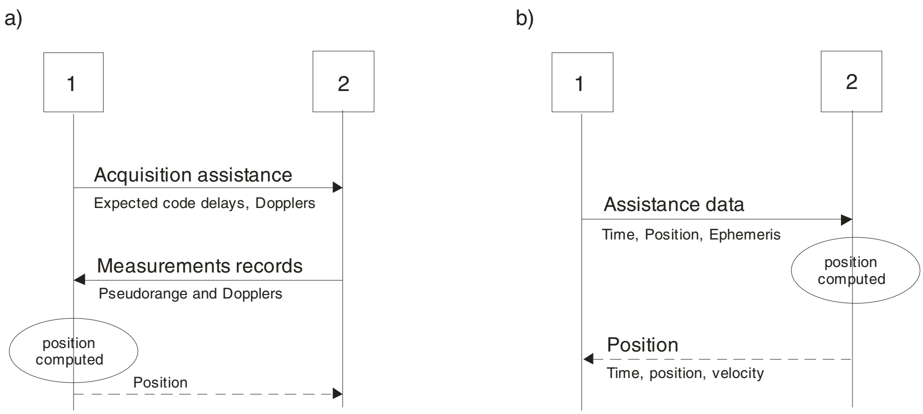


Fig. 1. A–GNSS structure: a) Mobile Station (MS) – Assisted, b) MS – based, 1 – Server and Base Station, 2 – Mobile Station [9]

In the MS–assisted method the position of the mobile device is computed in the network. This method acquires GPS satellite signals, makes measurements by correlating the locally generated PRN codes with the received GPS signals, and determines time-stamped pseudoranges. These results are transmitted to the A–GPS server, which performs the calculation of the MS position. The server provides the expected code delays and Doppler shift values. The MS–assisted solution can be adopted for tracking or navigation applications, only taking more signaling into account.

In the MS–based method, the mobile device (MS) computes its own position either with or without the assistance data and optionally returns the position solution to the location server if the location request originally came from the network. The MS incorporates a fully functional GPS receiver. The server provides the ephemeris and inospheric model, initial time and position. This data is used to compute the expected code delays and Doppler shift values, and additionally, after acquiring satellites, to compute the position. This method can be chosen for tracking or navigation applications because the assistance MSs are valid for 2–4 hours or up to 12 hours at the MS if the ephemeris life extension feature is used. Therefore the MS–based handset may work in an autonomous mode as well, providing position solutions to the user or embedded applications without the cellular network provided aiding data. The MS–based solution can be differentially corrected if DGPS corrections are sent to the handset, but the DGPS is not implemented in most current MS–based networks [9], [16].

5 Conclusions

- An Assisted –GPS can be defined as a system where outside sources, such as an assistance server and reference network, help a GPS receiver to perform the tasks required to make range measurements and position solutions.
- An Assisted–GPS technology offers significant performance advantages over either stand-alone GPS or mobile–station–based, particularly at low power levels often associated with consumer applications, such as the accuracy, availability, and coverage at a reasonable cost.
- The provision of assistance data provided by e.g. a mobile phone on a cellular network to a GNSS receiver greatly improves the Time To First Fix (TTFF) and increases the sensitivity of the GNSS receiver.
- The performance of any A–GNSS system is greater than of the same receiver in stand-alone mode because with the assistance from the network the receiver can operate more quickly and efficiently than it would do unassisted.
- As an A–GNSS receiver in the handset can detect and demodulate satellite signals, that are an order of magnitude weaker than those required by conventional GPS receivers, additional pseudorange measurements are available to it which can result in an increased positioning accuracy. Other advantages are a greater sensitivity and customer satisfaction.
- An MS–based solution is often more accurate than an MS–assisted, since the navigation software in the device can use measures such as post-fit residuals and the HDOP coefficient to estimate the quality of the position before completing a fix.

References

1. Januszewski, J.: GPS and other satellite navigation systems in urban transport. In: International Conference on Clean, Efficient & Urban Transport, CESURA 2003, Gdansk/Jurata (2003)
2. Januszewski, J.: Visibility of Satellite Navigation Systems in Urban Area. Artificial Satellites 38(2), 55–65 (2003)
3. Januszewski, J.: Geometry and Visibility of Satellite Navigation Systems in Restricted Area, Institute of Navigation, National Technical Meeting, San Diego (CA), pp. 827–839 (2005)
4. Januszewski, J.: Visibility and Geometry of Combined Constellations GPS with Health in Question. In: GLONASS and Galileo, Institute of Navigation, International Technical Meeting, San Diego, CA, pp. 1082–1094 (2010)
5. Information on GPS navigation..., `http://www.gps-practice-and-fun.com`
6. Djuknic, G.M., Richton, R.E.: Geolocation and Assisted GPS. Computer 34(2), 123–125 (2001)
7. Groves, P.D.: Principles of GNSS, Inertial, and multisensor integrated navigation systems. Artech House, Boston (2008)
8. Prasad, R., Ruggieri, M.: Applied Satellite Navigation Using GPS, Galileo, and Augmentation Systems. Artech House, Boston (2005)
9. Van Diggelen, F.: A–GPS, Assisted GPS, GNSS, and SBAS. Artech House, Boston (2009)

10. Gleason, S., Gebre–Egziabher, D.: GNSS Applications and Methods. Artech House, Boston (2009)
11. GPS World magazine, `http://www.gpsworld.com`
12. `http://www.navigadget.com`
13. British telecommunications company Spirent Communications plc, `http://www.spirent.com`
14. Samana, N.: Global Positioning Technologies and Performance. John Wiley & Sons, New Jersey (2008)
15. Kaplan, E.D., Hegarty, C.J.: Understanding GPS Principles and Applications. Artech House, Boston (2006)
16. Bryant, R.: Assited GPS, Using Cellular Telephone Networks for GPS Anywhere. GPS World 16(5), 40–45 (2005)

Using Databases in Switch Point Mechanism Diagnostics

Jakub Młyńczak

Silesian University of Technology, Faculty of Transport,
8 Krasinskiego str, 40-019 Katowice, Poland
jakub.mlynczak@polsl.pl

Abstract. This article presents an example showing the use of a database system in switch point mechanism diagnostic process. A switch point mechanism is one of the most important devices in railroad traffic control, responsible for the safety of trains passing through a switch. Up until now, software supplied with measuring devices has been used mainly to read and print the measurements made by the device. This article presents a concept of comprehensive analysis of a switch motor-switch system using a database system.

Keywords: Database, diagnostics, switch point mechanism, switch, diagnostic software, result analysis.

1 Introduction

Proper analysis of the condition of the switch point mechanism-switch system is a huge technical issue. On the one hand, we are dealing with switch point mechanism that are cared for by railroad automation technicians, while on the other hand, we are using switches cared for by the road maintenance personnel. Both sides have different goals. Automation technicians' primary concern is to keep the switch properly maintained with the lowest possible switching resistance and switch point spring-back. The road maintenance is merely intended to be capable of changing the switches without going into the switching resistance values. It is often the case that a switch with road legal parameters does not meet the requirements for switch motors. Moreover, the issue of measuring devices and specialised software used in keeping switch motors in good technical condition is a large problem in Poland (and elsewhere).

A proper measuring device is an indispensable device for ensuring correct parameters of the system. It should facilitate the measuring and analysing of forces in the switch point mechanism-switch system. It is used to measure the setting force, setting resistance as well as the holding force (with an external holding force source). It can be used to measure the impact of the switch on switch point mechanism when the train passes as well as the impact of switch points on the switch motor in set state.

The measuring device is, however, not all. Measurement is one thing, but data analysis is another. Specialised software allowing for the analysis of measurement results was, until recently, unavailable. Only recently, specialised software using database structures to store and analyse measurement results was released to be used in diagnosing the switch point mechanism-switch system [2]. Studies [4] and [5] have presented the issues of switch point mechanism-switch system diagnostics. This article discusses the concepts and structure of the software used in this process.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 152–159, 2010.

2 Software

The [2] application is a database system. It creates a database of switches, switch point mechanism and stations. It also stores information concerning measuring devices and users.

This information is stored in a correct format and exported to the measuring device.

The measurement results are imported directly back to the software from the instrument (via a USB connection) or from a different medium (most often a USB memory stick).

The software has three authorisation levels. The third and highest level is meant for a person, whose task is to assign authorisation levels to other users and who is able to delete the database. The second level has full authorisation to use the software, except for assigning authorisation levels and deleting the database. The first level is not authorised to delete any measurements. Levels 1 and 2 are used in standard operation.

The menu consists of three tabs: “Recording”, “HZM device” and “Measurements”.

The “Recording” tab allows the user to define stations, routes, switches and switch motors. Measurement devices and pins are real-time updates when the measuring device is connected. Moreover, this menu allows for creating motor and switch sets.

Additionally, the third level allows for specifying users and cleaning the database. Database cleaning cannot be undone. This, however, does not delete the specified types of switch point mechanism and switches. The database can be backed up, securing it against irreversible data loss data (e.g. after an accidental database cleaning).

The “HZM device” tab allows for connecting to a measuring device. This connection allows for exporting the database to the device and importing the measurement results back to the software. No instrument settings or user authorisations can be changed by the software.

In the “Measurements” tab, single measurement results can be printed and deleted. Moreover, when the measured object is not registered in the database (single objects or groups), the measurements can be assigned to specified objects.

2.1 Measurement Review

Measurements can be reviewed by selecting a station and switch from the menu on the left-hand side of the screen. The middle menu contains measurement dates and types. The setting and manual setting forces are shown in the cycle at the top, while the lower graph shows the setting resistance.

By selecting the measurements, they are displayed in the graphs. All measurement results are displayed in the same direction, regardless of the setting direction. Such approach allows for more precise measurement analysis, especially of the maximum force values. The software (as well as the device) shows the values of the setting force, setting resistance and trailable force.

The setting direction has been marked as follows:

> - left;
< - right.

The extreme values shown in the graphs correspond to the preset minimum and maximum values for given parameter. The illustration below shows the screen.

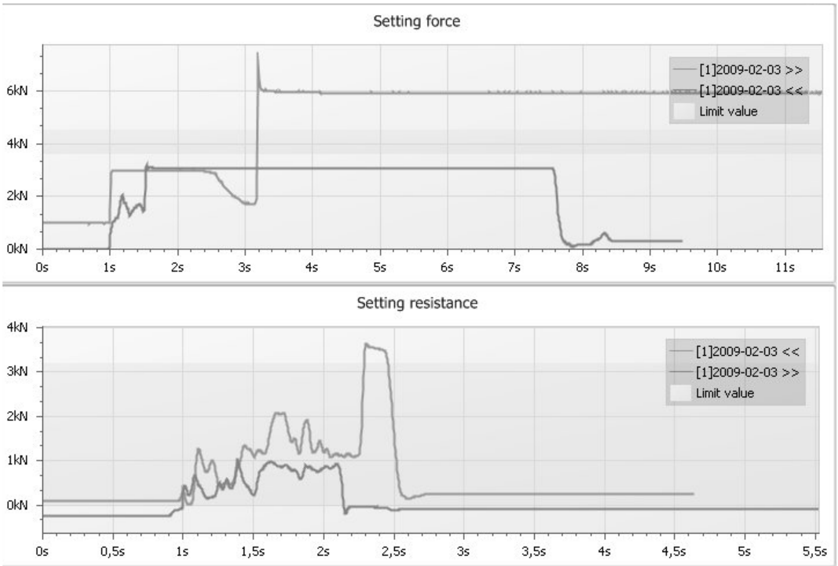


Fig. 1. Force characteristic curve with extreme regions marked

Measurement analysis is further facilitated by the ability to read the precise force value at a specified moment during the measurement. This can be done by hovering your mouse cursor for a specified moment during the measurement.

The software contains a predefined database of switch point mechanism models with their maintenance data. The user can then define the switch point mechanism model and technical data as well as modify the settings. The software automatically calculates the setting resistance value, which can be changed to a value recommended by the vendor (if it exists).

After a switch point mechanism is entered into the database, it shows up in the "uninstalled" group. This makes it easier for the database creators as they are not required to look for the new object among the already assigned ones.

The software can print study reports. Two print-out variants are possible. The first one is an automatic report. The software always selects the first measurement in a given category and direction (regardless of the number of measurements on a given day) and generates a report for these values. In the second type, the user specifies the measurements to be included in a report themselves.

2.2 The "Measurements" Tab

The "Measurements" tab is a place, where all measurements are stored with data concerning measuring devices, rods and measuring personnel.

If the measurement results are assigned to an object in the database, we can review, print or delete the measurements.

The tab becomes more functional when the measurements are not assigned to a given object in the database.

The standard functions presented here are only part of the system's capabilities. One of the advantages of the system is the ability to analyse the measurements and to compare the current condition to earlier readings. The result of such comparison is presented below.

The graph shows the readings for the setting force at switch no. 5 in Opole Zachod station. The measurements were taken on July 20, 2009 (darker) and August 24, 2009 (lighter). The graphs show that the July 20 measurements registered the setting force of 5.23 kN and that these measurements were made before the switch motor stopped. The August 24 measurements registered the setting force of 5.59 kN and were carried out by an operator, who finished the setting before the measurement and turned off the motor. Moreover, the graphs show that in both cases, the point tension is less than 1 kN (approx. 0.6 and 0.9 kN). These values are quite high and we should note that they work on the switch motor during the time when it theoretically should not be loaded by any force from the switch, as the setting closes the points. It can also be noticed that an obstacle of similar dimensions were used, as the time of coupling slipping is the same in both cases. It also interesting to add that practically every time the setting force is measured, there is a huge spike in the force when the overload coupling starts slipping. This is caused by violent interaction between the drive-train elements. It should be added that most devices measuring the setting force as well as most other result-viewing applications would show this spike as the setting force, which is a technical error.

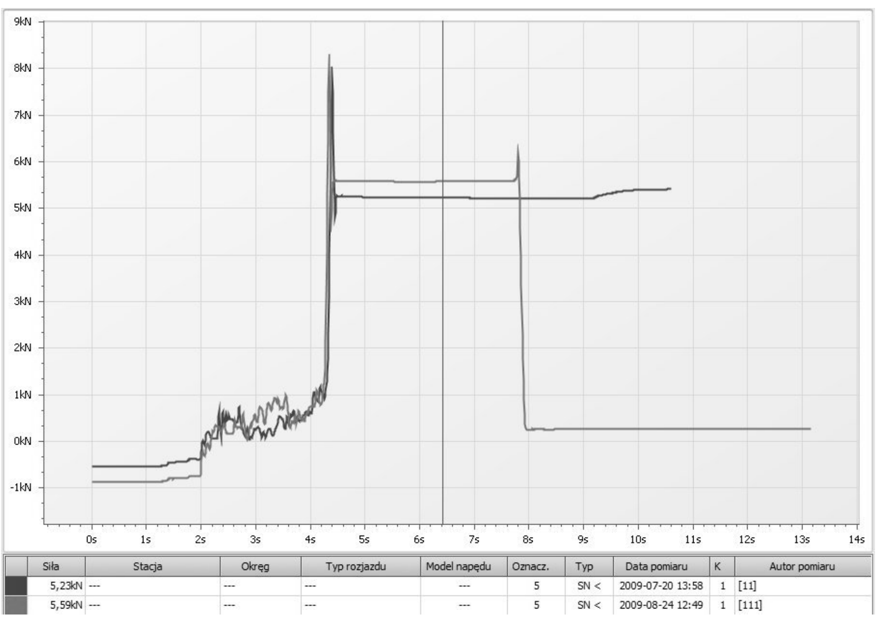


Fig. 2. Force analysis window

3 Database Structure

In order to conduct the diagnostic data analysis efficiently, a database has to be created. The database created for this software is an innovative solution for devices measuring the railroad traffic control. As can be seen in the general chart, the database structure is complex. This is caused by the fact that the software is used not only to review, but also to analyse the measurements. In order for the analysis to be complete and accurate, result layering should be feasible to analyse the results obtained at different times and in different atmospheric conditions. To facilitate decision-making, the measurement has to be assigned to a switch and switch point mechanism. Only then can the result be analysed, and further decisions concerning maintenance, repairs and inspections can be made.

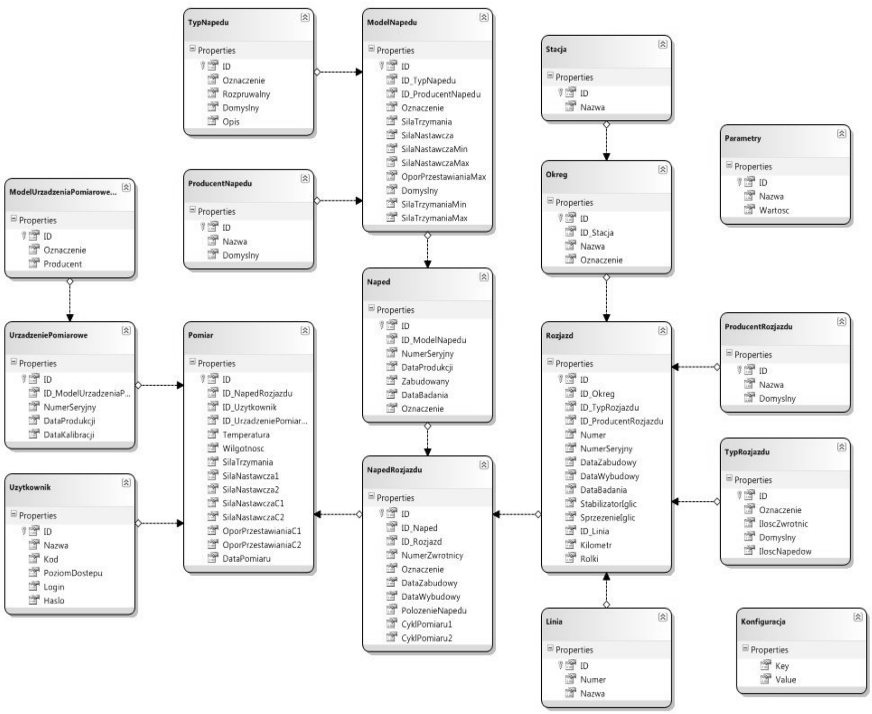


Fig. 3. General structure of the database used in the software

Fig. 3 shows the database structure used in the diagnostic software. It can be noticed that the whole base consists of 17 tables. The data flow between those tables is precisely defined. In order to precisely and unambiguously describe the parameters of the measured objects as well as all the parameters of the measurements themselves, many lines often have to be written about these objects. Figure 4 shows the structure of three tables in the database, i.e. the tables responsible for the measurement, the switch and the switch motor.

The measurement is described with fourteen lines that are required to describe unambiguously the measurement made by a diagnostic instrument. In order for the measurement to be unambiguous, information concerning the user ID (measuring person), atmospheric conditions and time of measurement should be stored and categorised, apart from the measurement results. The user ID is an element, which assigns a given measurement to the specified developer. Such a solution allows for “forcing” the employee to take accurate measurements and is proof of taking a measurement attached to the reading. This is especially significant when a dangerous situation happens (a railroad incident or crash). Information concerning weather conditions (temperature and humidity) is used in the comprehensive analysis of measurement results at different times and from different devices, so they can be analysed for optimising the device control processes. The date and time of measurement (assuming no modifications are made to the instruments), is an undisputable proof that the measurement was taken at a given time (as required by the railroad regulations and guidelines).

Fig. 4. Detailed structure of the “Measurement”, “Switch” and “Switch point mechanism” tables with their interrelations

The “Switch” table also requires storage and analysis of various data. Some of them are non-obligatory, but can be helpful in identifying an object (route number, km of route). The data concerning the date of installation (and possible removal of the installation) allow the condition of a device to be tracked by means of its location changes from the moment it is installed (new devices) until its technical death. Such an approach allows for the accurate analysis of the condition of a device over time as well as the condition of a switch/switch point mechanism, with which it operates.

The “Switch point mechanism” table, which describes the switch point mechanism, also contains information concerning the date of motor installation and removal of installation as well as the data of the switch where it operates (or operated – see “History” tab). The measurement times are also specified (the measurement cycles required by the regulations).

Apart from the database structure, the programming environment also allows for the generation of additional actions and information. One such example is the ability to create backup copies of the database stored in the software along with the measurements. Additionally, information can be added to an existing database (e.g. from regional services provided by a given team) containing data about certain regions.

```
public static class ZipUtil
{
  public static void ZipFiles(string inputFolderPath, string outputPathAndFile, string password)
  {
    ArrayList ar = GenerateFileList(inputFolderPath); // generate file list
    int TrimLength = (Directory.GetParent(inputFolderPath)).ToString().Length;
    // find number of chars to remove     // from orginal file path
    TrimLength += 1; //remove '\'
    FileStream ostream;
    byte[] obuffer;
    //string outPath = inputFolderPath + @"\" + outputPathAndFile;
    string outPath = outputPathAndFile;
    ZipOutputStream oZipStream = new ZipOutputStream(File.Create(outPath)); // create zip stream
    if (password != null && password != String.Empty)
      oZipStream.Password = password;
    oZipStream.SetLevel(9); // maximum compression
    ZipEntry oZipEntry;
    foreach (string Fil in ar) // for each file, generate a zipentry
    {
      oZipEntry = new ZipEntry(Fil.Remove(0, TrimLength));
      oZipStream.PutNextEntry(oZipEntry);

      if (!Fil.EndsWith(@"/")) // if a file ends with '/' its a directory
      {
        ostream = File.OpenRead(Fil);
        obuffer = new byte[ostream.Length];
        ostream.Read(obuffer, 0, obuffer.Length);
        oZipStream.Write(obuffer, 0, obuffer.Length);
      }
    }
    oZipStream.Finish();
    oZipStream.Close();
  }
```

Fig. 5. A fragment of source code responsible for archiving the database

4 Conclusions

The information presented above is an attempt to familiarise the reader with the trends in developing diagnostic methods for switch point mechanism - switch systems.

Due to the specificity of the whole system as well as to the high safety requirements, such diagnostics are required to meet a very high standard. There are currently a few devices like this in Poland and worldwide. However, the technical capabilities of many of these devices leave a lot to be desired. Diagnostic software is practically nonexistent. This is an important issue, as technical difficulties are often caused by the lack of understanding of a challenge. Measuring alone without analysis is insufficient, especially in Poland, where motor characteristics are stored in a paper form and filled out by hand with the obtained measurement results. Weather conditions and force fluctuations over time are unknown, not to mention any comparative analyses. The advisability of such an approach to the diagnostics of these systems may be illustrated by the applicability of this solution. Apart from European countries such as Poland, Germany, the Netherlands, Switzerland, Belgium and Ireland, such an application has also been for use in Hong-Kong.

References

1. Operating Instructions, Tongue-Force Measuring Device HZM for DB AG, Version: 1.0.1, Hanning&Kahl (2009)
2. User Manual – Diagnosis Software HZM System, Version: 1.2, Hanning&Kahl (2010)
3. Source code for HZMSystem Software, Version: 1.2, Hanning&Kahl (2010)
4. Mikulski, J., Młyńczak, J.: Cooperation of point machines with high speed switch points. In: 18 International Conference EURO-ŻEL 2010. Revitalisation of Economy - New Challenge for European Railways, Żylina (2010)
5. Nickel, H.: Data acquisition and diagnosis with the new HZM. TramNews 50, 12–13 (2009)
6. Białoń, A., Mikulski, J.: Wpływ typu ogrzewania rozjazdów na zużycie energii elektrycznej. Przegląd Elektrotechniczny 9, 37–39 (2009)
7. Luft, M., Szychta, E., Szychta, L.: Method of designing ZVS boost converter. In: Proceedings of the 13th International Power Electronic and Motion Control Conference, Poznań, pp. 478–482 (2008)
8. Surma, S.: Network Safety in railroad traffic control systems. Archives of Transport System Telematics 2, 41–44 (2009)

Baltic Ferry Transport

Janusz Uriasz

Institute of Marine Navigation, Maritime University of Szczecin,
Waly Chrobrego 1-2, 70-500 Szczecin, Poland
j.uriasz@am.szczecin.pl

Abstract. This article deals with the ferry transport in the Baltic Sea. Particular attention has been paid to shipping lines and ferries plying on them. Available ferries, their parameters, changes of basic parameters in time and functions of newbuildings are discussed. Characteristics of all the Baltic ferries and shipping lines are presented, according to data available in 2010.

Keywords: Baltic ferries, sea transport, sea highway, short sea shipping.

1 Introduction

Maritime transport is one of the most important components of the global economy. With its major contribution to development, it provides for the security and efficiency of other sectors of the economy. It should be underlined that maritime transport accounts for the carriage of more than 90% of cargo in the global trade. During the inauguration of the information campaign 'Go to Sea' in 2009, Secretary General of the International Maritime Organization, Efthimios E. Mitropoulos said 'when you look at the global transport of bulk cargo, grain and crude oil in particular, it follows that without maritime transport one half of the population would suffer from hunger, the other half from cold". Nearly 100 000 merchant ships are engaged in cargo transport.

In Europe approximately 40% of trade goes by sea [2]. The European Union takes action to increase this percentage and move as much as possible of the trade from land-based transport (about. 45%) to sea lines. The concept of Short Sea Shipping was first recommended by the Marine Industrial Forum in October 1992 on the promotion of short sea shipping and maritime multimodal transport. Guidelines for the development of sea transport in Europe are included in the White Paper issued by the European Commission 'European Transport Policy till 2010: time for decisions".

The short sea shipping concept is directly connected with another major concept of sea highways. The concept envisages improvement of the existing sea routes as well as establishing new regular ferry links between EU countries. In this way transEuropean network of highways will be created – natural extensions of land transport routes.

The involvement of Baltic states, including Poland, in the establishment of sea highways, will significantly influence the character and capabilities of the Baltic ferry shipping. The actual requirements will enforce putting new ferries into service or the modernization of the existing ones. In addition, port infrastructure will have to be expanded or modernized. The relevant work is already in progress or being planned in many ports, such as Świnoujście, where new ferry berths are under construction and berth No 6 is to be extended. The port of Gdynia has plans to build a new ferry terminal.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 160–167, 2010.

Sea highways are partly a result of the idea to build corridors spanning main sea areas in Europe. One of them is the North-South Axis project aimed at developing a concept of the transport corridor linking the Baltic and Adriatic Seas [5]. In Brussels in October 2009 14 regions of Poland, the Czech Republic, Slovakia, Austria and Italy signed a declaration of creating and developing a railway transport corridor.

Obviously, sea transport is not only concerned with cargo trade. As part of the transport chain, vessels also carry passengers, cars and trucks and trains. That is where specialized craft come to service – sea-going ferries.

Further in the article the full characteristics of all the Baltic ferries and shipping lines will be presented, according to data available in 2010 [3] and [4].

2 Characteristics of Sea Ferries

Sea ferries operate on regular lines between seaports. Their design and parameters depend on the port of their intended functions, trading area, route, specifics of harbours and actual berths they call at. Basically, they are divided into passenger, car, rail, passenger-car and passenger-car-rail ferries. Originally, ferries were designed for one type of cargo, e.g. cars. At present a majority of new ferries trading across the sea have more functions, e.g. they carry both ro-ro cargo and passengers. In such cases loading/unloading and embarking/ disembarking operations take place simultaneously through separate doors. Passengers generally get on board through passages directly connected with the terminal, while wheeled cargo goes through bow, stern or side doors. The design and equipment of modern ferries are aimed at increasing their functionality and passenger comfort. Ferries launched at shipyards are increasingly larger, which permits to accommodate more passengers as well as offer them a wider spectrum of hotel and food-serving facilities: cabins, restaurants, cinemas, theatres etc. Access to basic services such as the Internet, mobile communications or banking on board a ferry has become a standard. Passengers are able to communicate with their place of residence or work.

Changes take place in the size and functions as well as speeds developed by ferries. The latter divides ferries into the following groups:

- simple ferry sailing up to 15 knots,
- conventional Ro-Ro ferry – 15 to 23 knots,
- fast Ro-Pax – 23 to 30 knots,
- superfast Ro-Pax – 30 to 40 knots,
- HSC (High Speed Craft) – over 40 knots, mostly catamarans.

Traditional sea-going ferries are propelled by diesel engines, with the oldest ones still using heavy fuel oil. For environmental reasons – restrictions on exhaust gas contents of sulphur and carbon compounds emitted to the atmosphere lead to seeking new solutions. One such solution is electric motors fed with electricity from generators propelled by LNG. Gas-propelled engines emit 80% less contaminants than a traditional diesel engine.

3 Ferry Routes in the Baltic Sea

The Baltic Sea is an integral part of the European transport chain. Ferries operate in Europe in all sea areas as well as inland waters. The passenger and cargo throughput in inland river and lake waterways reaches significant levels. For instance, passenger traffic between France and Switzerland in 2008 amounted to 1.79 million people [3]. Apart from inland shipping, there are three main regions of ferry operation in Europe:

- Baltic Sea,
- North Sea,
- Mediterranean Sea.

The European ferry market is large indeed. In 2008 the total throughput (Fig. 1) was 761.5m passengers, 149.1m cars, 28.3m trucks. All the ferries went on nearly 5.1m voyages.

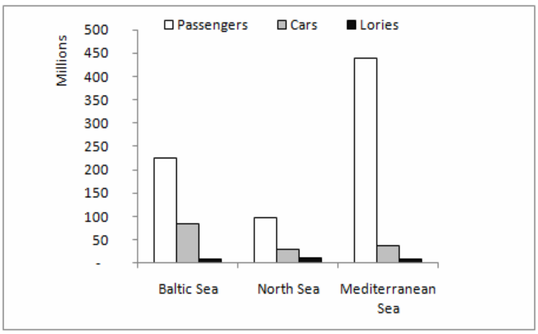


Fig. 1. Passenger transport by ferries by main European areas

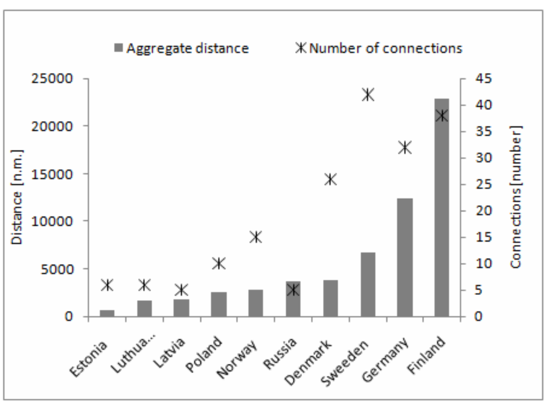


Fig. 2. Total international route lengths by Baltic countries

In the Baltic Sea there are total of 624 national and international connections (routes) are in operation, with some of them served by more than one carrier (shipowner, operator). The total length of all ferry connections is as long as 140,550 nautical miles (about 260,000 km), with the average route of 227.9 nautical miles. The longest lines run from the eastern corner of the Baltic Sea to western and southern Europe, over 1600 nautical miles long (the longest one links Kotka in Finland and Santander in Spain – 1 845 Nm). The Baltic Sea is crisscrossed by 97 international routes between 26 country combinations. Scandinavian countries, Finland-Sweden or Denmark-Norway, take the lead in these connections.

There are 42 routes to/from Sweden, 38 to/from Finland, while Poland has 10 routes [1]. The respective total lengths are: 6690, 22898 and 2565 Nm. Total length of routes leading to/from particular county is shown on Fig. 2.

The mean lengths of international ferry shipping by countries are presented in Figure 3 below.

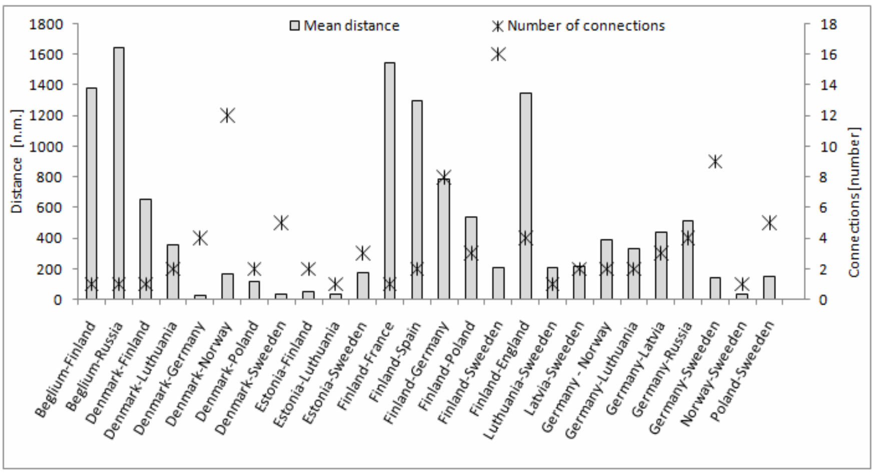


Fig. 3. International ferry routes in the Baltic Sea – mean route lengths

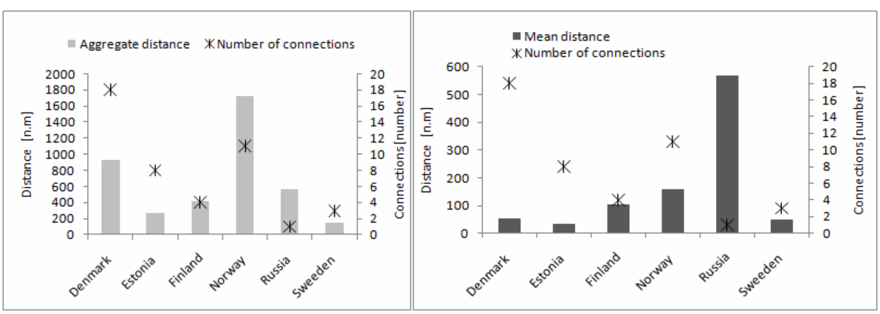


Fig. 4. Domestic routes in the Baltic Sea – total and mean route lengths

Naturally, the longest routes extend outside the Baltic, to western Europe, while those within the Baltic connect its eastern and western coasts. The mean Baltic route has a length of 280 Nm.

The 45 domestic routes in the Baltic Sea have a total length of 4000 Nm. These are mostly ferry services between Danish and Norwegian ports (Fig. 4).

4 Baltic Ferries

The density of ferry lines in Europe is the highest on the Baltic Sea. 202 ferries are operated on 142 lines. Despite the economic crisis, the Baltic ferry market is stable. Figures remain at a high level – about 200 million passengers a year, most of them between Scandinavian countries [3].

Of all the ferries sailing between Baltic ports, there are 13 HSC developing on average 37.5 knots with an average of 615 passengers on board. The brightest star in this group, Villum Clausen, can reach a speed of nearly 48 knots and embark 1037 passengers.

Other parameters in the analysis include:

- length,
- speed,
- age,
- passenger capacity,
- length of cargo line,
- gross tonnage,
- main engine power.

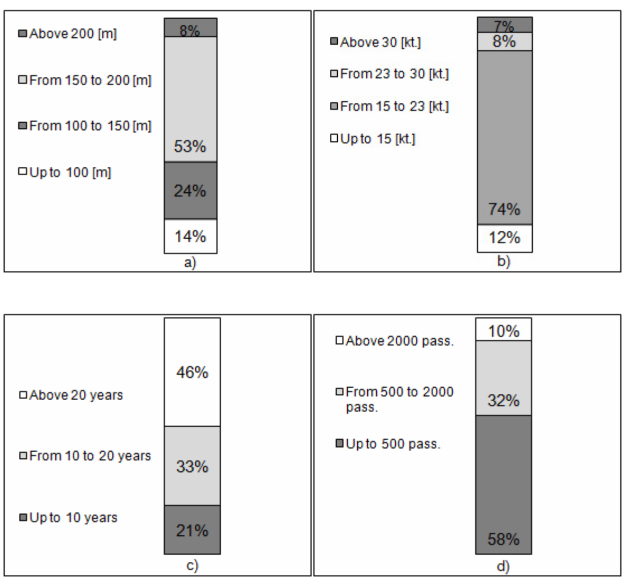


Fig. 5. Baltic ferries breakdown by: a) length, b) speed, c) age, d) passenger capacity

The largest ferries have a maximum length of 223.7 m (m/f Color Magic and m/f Color Fantasy). Ferries longer than 200 metres account for 8% of the total number (Fig. 5). Seventy three percent of the ferries develop speeds classifying them as conventional ferries, while 15% belong to fast or superfast vessels. Maximum speeds (excluding HSC) are up to 27 knots. Almost half of them are over 20 years old. One was built in 1953, all the others within this group were launched in the 1970s and 1980s. With the carrying capacity of the largest ferries exceeding 3,000 passengers. Ten percent or 18 ferries, make up a group of ships able to accept 2000 or more passengers each.

How many Ro-Ro cargo goes into the cargo decks depends directly on ship's cargo line length. This runs up to 4,200 metres, while 19 ferries boast the cargo line longer than 3000 metres. As shown in Figure Fig. 6a, 30% of the ferries have a cargo line length above 2000 metres. As far as tonnage is concerned, some ships exceed 75,000 gross tons. As many as 45 ferries have GT above 30,000 tons – 21% of the overall number of ships. All of them have powerful propulsion systems, with ME output of more than 30,000 HP each. The most powerful propulsion exceeds 68,000 HP!

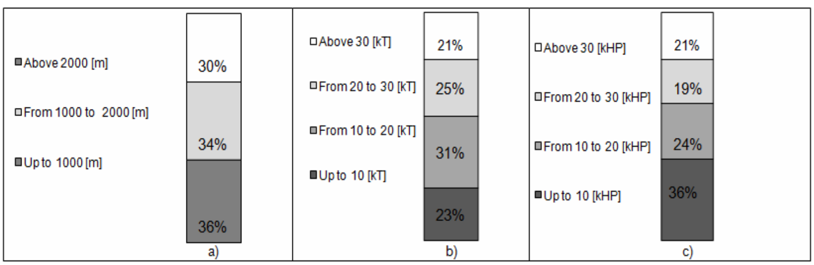


Fig. 6. Baltic ferries by: a) length of cargo line, b) gross tonnage, c) main engine power

Mean parameters of Baltic Ferry are presented in table no 1.

Table 1. Mean parameters of a Baltic ferry

Age	Length	Breadth	Draught	GT	Speed	Passenger	Cargo line
18,79	156,95	24,00	6,01	21 798,79	20,36	802,43	1764,89

Today, ferries sailing between Baltic ports are 19 years old on average. In 2010 thirty eight new vessels were put in operation, of which 22 are Ro-Pax vessels carrying cars and passengers. Nine of them develop a speed over 23 knots. In retrospect we can see (Fig. 7) that ferries built in the past were slower, reaching service speeds between 15 and 20 knots. New ferries break speed records, no wonder the general trend is rising. Excluding HSC, the graph below shows that newbuildings are on average 5 to 7 knots faster than the oldest craft.

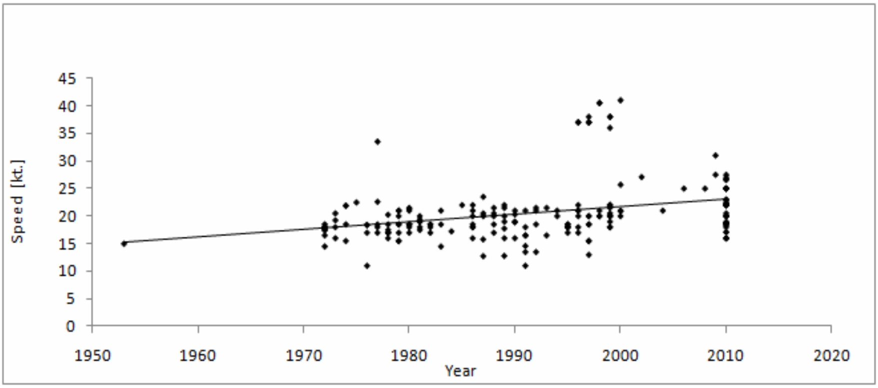


Fig. 7. Speeds of new ferries put into service on the Baltic

Rising trends in transport capacity of newly built ferries are also reflected respectively in their lengths and, consequently, cargo line lengths. The growing trend seen in diagram above well illustrates characteristic changes of Baltic ferries.

5 Conclusions

The history of Baltic ferry construction dates back to the 1920s. The ferry is understood here as a ship capable of carrying passengers (with special spaces for them), cars (on decks, including open decks) and other wheeled vehicles, sailing on a regular line. After numerous changes in the type, size, functions and manoeuvring characteristics, the present design trend has led to Ro-Pax ferries of increasingly higher speeds – Fast and Super Fast. They accommodate a lot of passengers and a lot of wheeled cargo. As the tendency to build larger and faster ferries continues, we are likely to see a ferry with the cargo line over 7,000 metres.

Baltic ferries will be an essential link of transport chains connecting Baltic states, as well as Mediterranean ports with those in the north of Europe. They make up an indispensable part for the effective implementation of transport corridors within the TEN-T concept.

The functionality of ferries, including places to lie down, depends on route length, or shall we say, voyage time.

- short routes – ferries sail at daytime only; such vessels generally do not have hotel/restaurant infrastructure;
- medium length routes – voyage time is about six hours; those ferries mostly sail at daytime, although part of a trip may take place in the night, so sleeping spaces have to be offered;
- long routes – voyage time exceeds 12 hours, most often at night time; those ferries have to offer hotel/restaurant facilities.

All three types of ferry line exist on the Baltic. Those running to and from Poland are medium-length and long ones. If spanned by high speed craft, a number of short routes might be established. However, HSC are restricted by weather, so they could only be operated in summer.

Nowadays designers changed their approach towards comfort and attractions of the ship [6]. Good food and opportunities to relax have become priorities. As in the past a voyage by ferry was regarded as a waste of time, the prevailing motto was: the faster, the better. Technological advancements (information technologies) enable passengers today to spend time effectively at any time and place, a sea-going ferry included. Consequently, one day we may see another change on the Baltic: from a Ro-Pax to a Ro-Cruise!

References

1. Bialas-Motyl, A.: Selected information on sea transport collected in accordance with the requirements of 95/64/WE Directive. In: 6th International Scientific Conference — Marine Ports, pp. 1–14. AM Szczecin, Szczecin (2006) (in Polish)
2. Sanaba, M.: Short Sea Shipping efficiency analysis considering high speed craft as an alternative to road transport in SW Europe, Phd Thesis, Univeristat Politecnica de Catalunya (2009)
3. Market:09 Statistics, Market reports & outlook for ferry, cruise, ro-ro and hi-speed shipping, ShipPax Information, Halmstadt (2009)
4. Sea-web Lloyd's Register database, `http://www.sea-web.com`
5. South-North Axis project, `http://www.sonoraproject.eu`
6. Uriasz, J.: Characteristics of ferries operating in Baltic Sea. In: 6th Sea Forum, Kolobrzeg, pp. 233–245 (2007) (in Polish)

Estimation of the Mean Velocity of a Group of Vehicles in Strong Random Noise Conditions

Jan Purczyński

Higher School of Economics and Technology,
Klonowica 14, 71-244 Szczecin, Poland
janpurczynski@onet.eu

Abstract. A method of estimating the mean velocity of a group of vehicles moving in one lane was proposed. The vehicle image blurring and the appearance of Gaussian noises at individual stages of image acquisition were considered. The method is based on the signal approximation by application of a parabolic function (formulas (1) and (2)). The algorithm work in strong random noise conditions was verified through appropriate computer simulations.

Keywords: Vehicle velocity estimation, random image distortion, computer simulations.

1 Introduction

In this paper the problem of vehicles velocity estimation based on the image registered by a video camera was considered. In the case of low level noise, the method based on subtracting subsequent image frames is used, which enables identification of moving objects. One such example is an algorithm proposed by Lucas and Kanade [1]. One of the versions of this algorithm includes an application of image derivative. However, signal differentiation procedures in the presence of random noises may yield large errors [3]. In the case of higher noise, the method of subtracting subsequent image sequences becomes useless, since it leads to an increase in the noise level – a summation of noise variance occurs.

In this paper, in which strong random noises are considered, it has been assumed that for individual frames an object position is estimated, and that the velocity is determined on the basis of image sequences (10 or 20). The method will be illustrated using a single moving vehicle, and subsequently, it will be implemented to estimate the mean velocity of a group of vehicles.

2 Estimation of a Single Vehicle Velocity

It is assumed that before the velocity measurement started, a sufficient number of background images had been registered (e.g. 1000), on the basis of which averaging was done, which enabled identification of characteristic background features. It is also assumed that an averaged background image is subtracted from subsequent registered images with a moving vehicle. As a result, it is assumed that the background has a level equal to zero and is disturbed by a Gaussian noise in the same way as was the

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 168–174, 2010.

vehicle. In order to simplify these deliberations, vehicles in a simplified grayscale are considered: a figure AS=0.5 is assigned to a light-color vehicle, and a figure AS=-0.5 is assigned to a dark-color one. Assuming that a car velocity equals 90 km/h and that a camera registers 25 frames per second, it can be found that two subsequent frames correspond to the distance of one meter covered by the vehicle. Assuming that the video camera has a resolution of 320 pixels and that the image covers 320 meters of the road, it can be concluded that the distance between two adjacent pixels of the image corresponds to the distance of 1 meter of the road. It is also assumed that the movement of a group of vehicles – each 6-meter long - takes place in one lane, and that the camera is placed on a side of the road.

A model of a vehicle, i.e. a rectangle with a base of 6 pixels, has been blurred using moving average filter, and has taken the shape of trapezium *ys* shown in Fig.1.

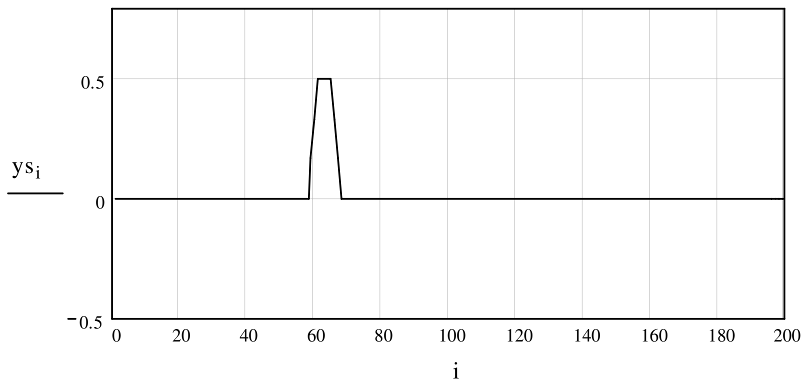


Fig. 1. Model of a single vehicle

The figure presented in Fig. 1 has been disturbed by a Gaussian noise with its standard variation $\sigma = 0{,}3$, which has been shown in Fig. 2.

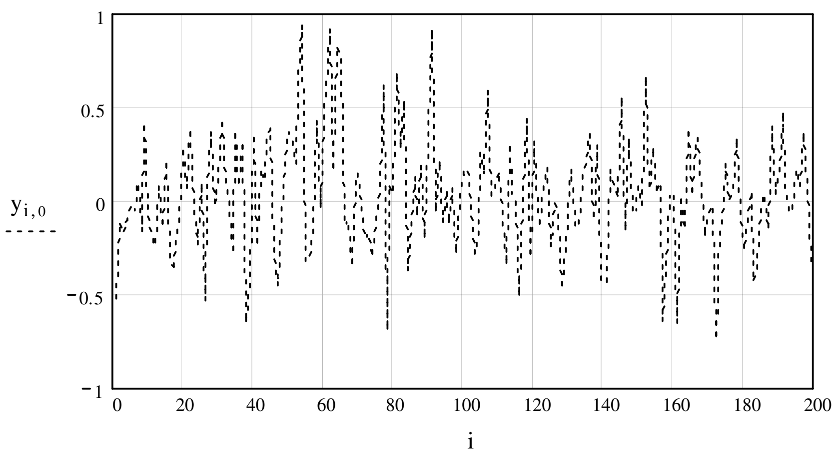


Fig. 2. A sample figure representing a vehicle in the presence of random noises obtained through computer simulations

In order to estimate a vehicle position, a parabolic approximation was adopted in the following form.

$$f(i, j) = a0_i + a1_i \cdot j + a2_i \cdot j^2 \tag{1}$$

where $j = -M, -M+1, .., M$.

Trend parameters were determined using formulas:

$$a1_i = \frac{\sum_{j=1}^{M} j \cdot \left(y_{i+j} - y_{i-j}\right)}{B}$$

$$a2_i = \frac{\sum_{j=1}^{M} \left(y_{i+j} - y_{i-j}\right)\left(j^2 - B\right) - B \cdot y_i}{D} \tag{2}$$

$$a0_i = \frac{\sum_{j=1}^{M} \left(y_{i+j} + y_{i-j}\right) + y_i - C \cdot a2_i}{2M+1}$$

where: $B = \frac{M(M+1)}{C}$; $C = B(2M+1)$; $D = \frac{C}{15}(2M-1)(2M+3)$

y_i - registered signal

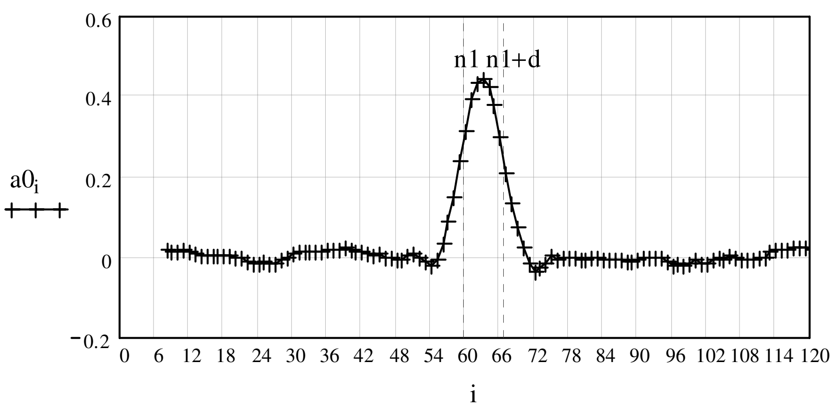


Fig. 3. Averaged values of coefficient $a0_i$. Vertical lines define the interval (n1, n1+d) in which a vehicle is present.

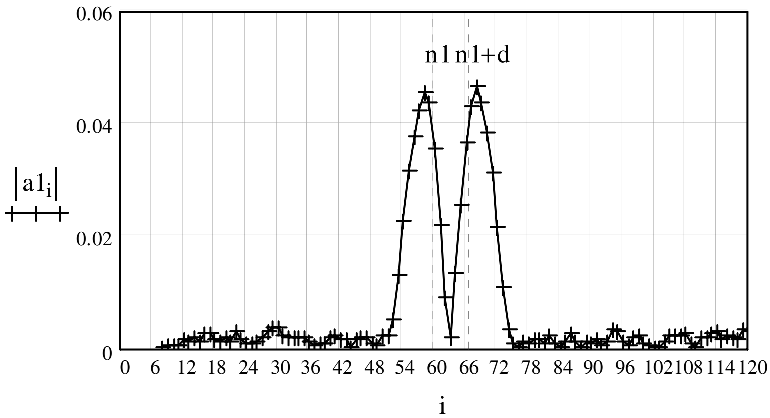


Fig. 4. Averaged absolute values of coefficient $a1_i$ .Vertical lines define the interval (n1, n1+d) in which a vehicle is present.

Fig. 3, Fig.4. and Fig. 5 show the values of trend parameters obtained by averaging the results of a hundred of computer simulations.

From these figures, three methods of vehicle center estimation are obtained:

- coefficient $a0_i$ maximum method
- $|a1_i|$ minimum method, having earlier determined the position of maximum values
- coefficient $a2_i$ minimum method.

As a result of a numerical experiment, the strongest resistance to random noise proved the method based on coefficient $a0_i$ maximum. In this paper this method was applied (marked as variant I) along with its modification which consisted in the search for the maximum value of the product L of subsequent coefficients $a0_i$ (variant II).

The algorithm of vehicle velocity estimation is the following. For consecutive (e.g. 10) images we determine the values of pixels for which the maximum of coefficient $a0_i$ occurred, as well as their median. The values of pixels whose distance from the median exceeds 20 are replaced with the mean value of their neighbors. In the case of outermost pixels (the first or the last), the linear extrapolation of the values of two closest pixels is performed. For the obtained set a linear trend is determined. Its slope of a straight line determines the vehicle velocity. In the case of variant I, root mean square error $rmse_{10} = 0{,}129$ was obtained for the sequence of 10 images and $rmse_{20} = 0{,}040$ for 20 images – calculations were made for a window 2M +1 = 15 (formula (2)). Variant II yielded the following error values: $rmse_{10} = 0{,}109$, $rmse_{20} = 0{,}033$ - for M = 7 and L = 4 respectively.

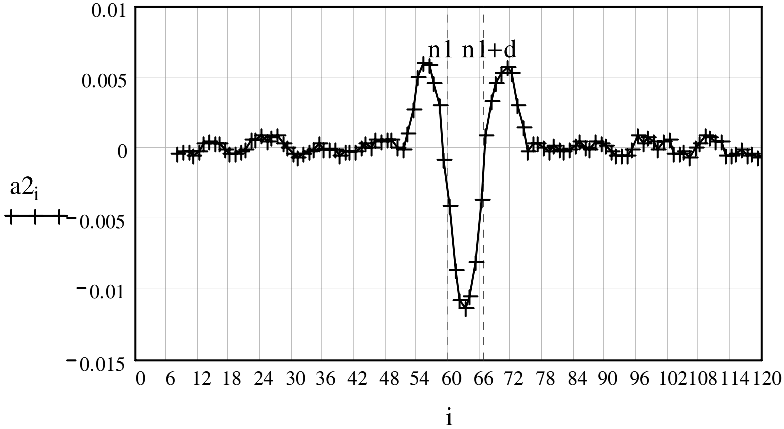


Fig. 5. Averaged values of coefficient $a2_i$.Vertical lines define the interval (n1, n1+d) in which a vehicle is present.

The limitation of this method is the number of wild pixels with the values diverging strongly from the median. In the case of noise of the level $\sigma = 0{,}3$ it constituted up to 30%, which guaranteed the right median value. For $\sigma = 0{,}4$ erroneous observations constituted up to 50%, which rendered correction impossible.

3 Estimation of Vehicle Group Velocity

Fig. 6 shows the model of a group of three vehicles – each 6 pixels long (6m) – with the distance between the centers of the vehicles equal 15 pixels (15m), which means that the real distance between the vehicles was 9 pixels (9m).

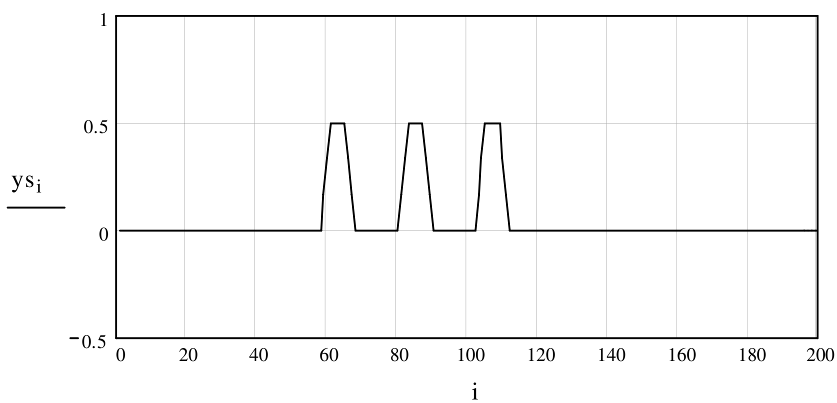


Fig. 6. Model of a group of vehicles

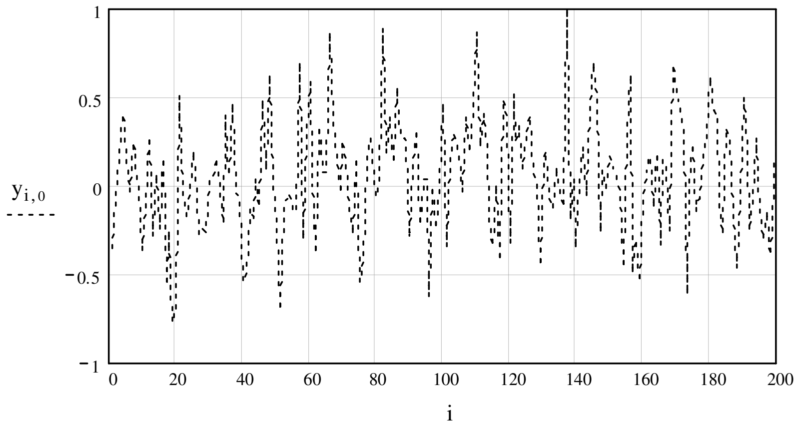


Fig. 7. Fig. 6 disturbed by Gaussian noise with standard deviation $\sigma = 0{,}3$

Fig. 7 shows an example of computer simulations in which random noise ($\sigma = 0{,}3$) was added to the image shown in Fig. 6.

The algorithm of velocity estimation for a group of vehicles is similar to the algorithm described in section 2. For a given image, pixel i_1 coordinates of coefficient $a0_i$ maximum are determined, and subsequently, this action is repeated excluding points $i_1 - 5, i_1 - 4, ..., i_1 + 5$, obtaining the value of pixel i_2. Another exclusion of points $i_2 - 5, i_2 - 4, ..., i_2 + 5$ allows to determine coordinate i_3 corresponding to the third extremum. The determined coordinates i_1, i_2, i_3 are then put in order of increasing value. The described actions are repeated for every image. Subsequently, three groups are determined: the first including the coordinates of pixels which, as a result of arranging, were placed in the first position; the second including pixels in midposition; the third including pixels in the last position. For each group, which represents the location of an individual vehicle, its velocity is determined – following the algorithm described in section 2.

Table 1 includes the results of computer simulations performed for the sequence of 10 images. In both variants (I and II) the window size (formula (2)) was 2M + 1 = 15,

Table 1. Results of vehicle group velocity estimation

Variant	Vehicle	Velocity	bias	rmse	mae
I	I	0,976	-0,024	0,130	0,106
	II	0,943	-0,057	0,130	0,108
	III	0,918	-0,083	0,133	0,115
	group	0,945	-0,055	0,086	0,074
II	I	1,017	0,17	0,074	0,055
	II	0,994	-0,006	0,037	0,029
	III	0,979	-0,021	0,087	0,067
	group	0,997	-0,003	0,040	0,031

however in the case of variant II, product L = 6 of consecutive values of coefficient $a0_i$ was considered.

The last column of the table includes the values of mean absolute error.

The most important error criterion is *rmse*, since it is a square root of mean square error (*mse*), where *mse* is the sum of variance and the square of *bias* [2].

The velocity of a group of vehicles was derived as a mean of results for individual vehicles obtained for subsequent computer simulations. Looking at the table, it can be concluded that the velocity of a group of vehicles in Variant II has smaller bias than velocity biases for individual vehicles. However, the value of error *rmse* became smaller for both Variants I and II.

4 Conclusions

In this paper we proposed a method of estimating average velocity of a group of vehicles based on the image registered by a video camera including high level noise. The algorithm is based on the approximation of a signal using a parabolic function (formulas (1) and (2)). Two variants of the method were considered: a direct search for the position of the maximum of coefficient $a0_i$ (variant I) as well as determining the position of the product maximum of L consecutive values of this coefficient (variant II). Having analyzed the data presented in Table 1, it can be concluded that Variant II yields half as large errors (rmse, mae) as Variant I, and that the average velocity of a group of vehicles is burdened with a smaller error than the errors of velocity values for individual vehicles.

In calculations also the cases of AS = - 0.5 (a dark vehicle) were considered. Their results were similar to those presented in Table 1.

Taking into consideration the fact that the calculations are made for the sequence of 10 images, which corresponds to the time of registration t = 0.4s, it can be concluded that the method concerns the instantaneous value of velocity.

Acknowledgements. This work is supported by the Polish Ministry of Science and Higher Education (Grant No. N509 399136 „Estimation of the vehicles' motion trajectories using the Bayesian analysis and digital image processing algorithms").

References

1. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceeding of Imaging Understanding Workshop, pp. 121–130 (1981)
2. Krzyśko, M.: Statystyka matematyczna. Cz.II. UAM, Poznań (1997)
3. Purczyński, J.: Algorithms for differentiation of signals with random noise. In: Computer Applications in Electrical Engineering, Poznań, pp. 21–31 (2009)

Using Telematics in Transport

Jerzy Mikulski

Silesian University of Technology, Faculty of Transport,
Krasinskiego 8, 40-019 Katowice, Poland
jerzy.mikulski@polsl.pl

Abstract. Design and development of telematics transport systems represents a new approach to solving transportation problems. Telematics offers large opportunities to strengthen the positive features of transport (availability, mobility), while minimizing its negative impacts (e.g. environmental pollution, energy consumption, congestion, accidents, infrastructure construction costs) without any additional spending on investment.

Keywords: Transportation, ITS, Telematics applications.

1 Introduction

As a technical term, “telematics" is a portmanteau of “telecommunication” and “informatics”, while as a science, it integrates research in telecommunication, automatics and information technologies. Currently, telematics is defined as telecommunication, information and IT solutions, as well as automation solutions adapted to the needs of physical systems (and often integrated into these systems).

Physical systems with such telematic solutions are called "Intelligent Systems". Intelligent systems use numerous devices (and software), such as:

- electronic communication systems,
- data acquisition systems,
- operator data presentation systems,
- user data presentation systems.

2 Applications of Telematics

Telematics is a modern field in technology. One can say that "telematics has become trendy recently". In recent years, telematics can be found in various areas of economy, often accompanied with a qualification of its applications: construction telematics, medical telematics, transport telematics, etc (see Table 1).

We are concerned with transport telematics, which encompasses all means of transport. Transport telematics means the application of teleinformatics technologies in managing transport systems. Various telematics systems apply to different types of transport infrastructure and to different means of transport. Therefore, telematics applies to all means of transport.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 175–182, 2010.

Table 1. Applications of telematics

Sector	Area	Application
Construction	institutions, industry	safety, security, lighting, transport, fire and access control
Power engineering	power supply, electricity, oil, gas, alternative fuels	turbines, generators, batteries, backup supplies, wind power plants fuel cells,
Household	infrastructure, safety, comfort	power, supply, washing machines, dishwashers, lighting, heating, ventilation, safety cameras, supervision
Medicine	healthcare, nursing, medical examinations, science	telemedicine, telesurgery, monitoring
Industry	resources, automation, distribution, science	manufacturing, engines, motors, pumps, valves, conveyors, processing, assembly, packaging,
Transport	means of transport, infrastructure, transport systems	vehicles, vessels, airplanes, signalling, navigation, signs, payments
Commerce	warehouses, deliveries	terminals, cash registers, ticket vending machines
Safety	supervision, tracking, equipment, public structures, crisis management	police, ambulances, helicopters, fire brigades, monitoring, internal safety
IT	public, industrial, personal	power supplies, UPS, routers, memory, switches

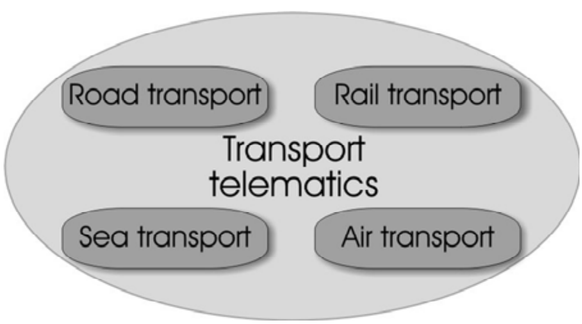


Fig. 1. The scope of transport telematics

In transport, telematics allows for influencing the transport process (both passenger and cargo) in order to improve the effectiveness, enhance planning and safety as well as to reduce the environmental impact. Telematic systems enable a more effective use of transport infrastructure, increase traffic quality and flow as well as facilitate transport systems operation. The ability to retrieve, process and transfer traffic data in real time provides new opportunities in mobility.

Table 2. Types of telematic systems

Applications	Services
Traffic management	Traffic control Law enforcement Planning support Incident and accident management Demand management Infrastructure maintenance Priority vehicles
Safety	Intelligent intersections Transport safety Safety of weaker vehicles
Public transport	Public transport management On-demand transport management Information concerning travel time Pre-travel information Information during the travel Personal information services
Means of transport	Guiding along a route Navigation Visibility enhancements Automated driving Collision avoidance Advanced monitoring systems Automated inspection Internal driving safety monitoring Vehicle fleet management
Support	Accident information and personal safety Rescue vehicle management Hazardous materials and accident notification
Electronic payments	Electronic financial operations

All over Europe, the implementation of intelligent transport systems (ITS) is performed according to plans and EU recommendations. In Poland, the process of implementing the telematic systems is slow and, thus, less beneficial than expected.

Transport telematics means the systems allow for influencing the behaviour of drivers, the operation of vehicle elements, or the route through information transfer and analysis.

2.1 Rail Applications

In railroad transport, traffic control applies to the gaps between trains, speed and security against entering wrong routes. Due to long braking distances, railroad traffic is controlled from an external control room, while the engine driver is informed about the traffic circumstances by light signals or in the cockpit. On-board instruments can take over when the driver disregards signalled situations (or is unable to drive).

Ensuring the safety of high-speed railroad traffic requires constant exchange of information between the rails and the vehicle. Operations are currently under way to implement the latest traffic control systems using new communication technologies for data transmission.

European railroads are trying to unite the widely diversified national traffic control systems according to ERTMS/ECTS project (European Rail Traffic Management System/European Train Control System), and to increase the throughput by abandoning the traditional system of linear barriers and implementing the movable distance system (virtual barriers), which is enabled by constant train tracking from the control centre, and transferring the extreme traffic parameters directly to the train. The system requires precise train positioning and train continuity tracking.

2.2 Road Applications

The purposes of telematic systems in road transport:

- Traffic control,
- Traffic flow control,
- Cargo flow tracking and optimisation,
- Fee collection,
- Travel information and booking.

The systems should therefore be able to control traffic lights and changes in traffic directions, regulate the merging process, influence driving speed and gaps between vehicles as well as warn against hazardous situations (such as congestions, surface damage, weather conditions, etc.).

In order to warn the drivers about hazardous situations, the dynamic content can be used in addition to static signs. In controlling the traffic en route, there is a problem of the drivers not obeying the limitations, even though such matters are regulated by the Highway Code. In the future, such information could be transferred directly to the vehicle. There are already systems to enforce a certain driving style using technical solutions (such as limiting the top speed or automatically keeping the distance from the preceding vehicle). Navigational information is especially important, although they should be accompanied by information concerning the real arrival time. Navigation systems are very popular, and static data concerning the roads (maps) allow for reliable routing, when combined with the GPS signal. Such navigation devices can provide dynamic information concerning traffic and detours over radio frequencies or cellular data connections. Personalised (dedicated) routing service is also possible. Current traffic information could also be provided via an Internet connection. Such system could be warning against long-term (e.g. roadworks) or temporary (e.g. traffic jams) obstacles.

2.3 Management Applications

Parking management systems allow for directing individual vehicles to a free spot using the shortest route. This can be a solution to the long search for free parking spots, and could allow for more even car park utilisation. The dynamic parking system becomes especially handy in rush hours.

The main task of the transport system operator is to control the traffic of municipal transport. Any disruptions, caused mainly by overloads on certain sections, are avoided owing to the dynamic transfer of information or commands. Public transport allows for route changes (by an operator), as well as for decisions made by the passenger themselves (changing their means of transport). By controlling the traffic flow in the streets, detours can be organised for overloaded or blocked sections. The throughput of certain sections of the road network can thus be enhanced by using alternative routes.

Public transport has been using telematic systems for the last several years to identify disruptions in scheduled runs (delays, disrupted connections with other routes) and to remove them or limit their effects. This applies to both rail and road urban means of transport.

2.4 Automated Charging Applications

There are already systems for automated charging in public transport. There is, however, an issue of specifying the correct fee. Currently, the system uses (apart from normal tickets and proximity cards) a "entry/exit" card, which records the moment of getting on and off the vehicle, and uses this information to calculate the rate. The system also considers any discounts for retired people, students and regular passengers. There is also a system, which uses the passengers' mobile phones to charge them. Both systems (recording passengers getting on and off at stops) provide information concerning the traffic and the direction, which used to be collected by surveying passengers. The knowledge of current passenger needs may lead to a wider adoption of flexible transport systems, e.g. on demand transit.

The ticket card systems allow for adopting various methods of serving passengers, such as creating different tickets for different routes, rate selection, but also using the card as a library card, parking card, cinema ticket, etc.

In road transport, just as in the railroads, any congestion can be overcome by directing the vehicles to a different route, which can have a vast significance in municipal traffic.

In the last couple of years, telematics has led to a vast improvement in quality and optimisation of transport, especially in freight transit. The most important element of the system is the online cargo and vehicle localisation, with the ability to compare the location to a planned route. This provides the customers with up-to-date information. It also allows for remote access to the data concerning the state of the goods carried, such as the temperature inside the vehicle and other similar parameters, and for notifying the sender in case any of the parameters are exceeded. This makes shipping hazardous materials far safer than before. The technical condition of a vehicle can also be controlled remotely.

Cargo can be tracked, giving full control over the whole shipping chain. Many of the opportunities provided by telematics give the road freight shipping an edge over other means of transport. The high quality achieved through telematics is becoming a standard among forwarders.

Telematic systems allow for paying for paid roads without the need to stop the vehicle. It requires an appropriate road infrastructure as well as the installation of special devices that would enable telematics charging. This applies to both commercial and personal transport. The charging systems calculate the fare according to the distance travelled. The fares are charged using telematic infrastructure and satellite navigation or radio communication.

2.5 Other Applications

Information concerning travel conditions should be available both before the departure and during the trip. Such information include a wide array of data: from the statistics concerning the road network for personal transport and public transport routes, schedules and fares, to dynamic data concerning road conditions and traffic (congestions, delays). Data used in statistical information originate from the infrastructure management systems.

Dynamic information for personal use are compiled based on reports concerning road conditions and traffic obtained from automated measurements, surface sensors, automated assessment of video feeds as well as reports sent directly from traffic vehicles. In public transport, traffic data come directly from the vehicles to the traffic control centres.

Such information concerning travel conditions can be accompanied by further tasks needed when travelling (e.g. ordering a taxi, luggage handling, hotel booking, booking tickets to cultural events, etc.), but it mainly applies to ordering transport tickets.

The potential of transport telematics to resolve traffic overloads and enhance passengers' comfort seems unquestionable. We should, however, bear in mind that there are certain factors that can hinder the development of transport telematics.

Telematic systems require costly technical components (such as measuring and information equipment, telecommunication devices, software). The operating costs of telematics consist mainly of equipment cost (hardware and software), the costs of accessing data and of providing the service. These costs are borne by both system operators and the end users. The benefits of transport telematics cannot be precisely specified until the influence and effects of all the systems are sufficiently known. The benefits to transport should be separated from benefits to the environment, the economy and the society.

When directing the traffic in individual sections as well as in the whole network, the conflict concerning the optimal telematic system arises. Telematic systems operators want all the participants to achieve the shortest possible trip time, the highest possible safety with low environmental impact. Therefore, they strive for the most evenly distributed traffic in order to avoid concentrated overloads. Individual drivers, on the other hand, care most about minimising the cost and time of travelling. A combination of these two goals can only be achieved when passenger behaviour becomes beneficial for the user when obeyed (e.g. using a non-optimal route would cause increased fares).

The potential of telematics in tracking the environmental impact of transport is vast. This may include routing with consideration to environmental limitation, using telematic information, charging and control systems. The systematic implementation of telematic technologies will lead to the inclusion of all transport systems, which may limit the dominating positions of road transport.

One of the biggest obstacles in further development of transport telematics is the technical integration of many different systems. This stems from fast innovation and the lack of standards in the field. In this situation, the search for common system architectures seems well-grounded.

3 Implementation and Legislation

Another problem faced by transport telematics is the duration of the implementation process. This time is often longer than the development cycle for the new technology. When such system is finally ready to be used in practice, it often turns out to be technologically outdated. For the sake of progress in telematics, we cannot abandon an affective technology, even though new solutions appear.

There are also certain legal issues hindering popular acceptance of telematics. Effective implementation of some telematic systems in transport requires people to obey certain norms of co-existence. A question of legal responsibility for personal data protection thus arises. Many transport telematics systems collect personal data from unknowing passengers. One such example may be a GSM phone, which constantly records its user's location. All GSM-based telematic technologies allows for tracking user’s location. The same applies to ticket purchases, etc. Combining such data with other information may afford the authorities, such as law enforcement officers, a far-reaching supervision of citizens. This is a violation of current personal data protection laws. This is why further development of telematics should use system architectures that provide a high level of anonymity to all concerned parties.

4 Conclusions

Individual transport telematic systems are currently at different stages of development. Today, we should concentrate on supporting the development of telematics to better meet our obligation to provide safe and unobstructed transport. Our R&D efforts should be tackling the problems of telematic systems' operation and integration.

As the technological possibilities keep expanding, telematics will quickly enter into all areas of our lives, especially transport. Telematic systems will become an increasingly indispensable element of transport. In order to provide the society with the most benefits, we have to recognise the positive influence of telematics as well as any risks and obstacles to its development.

Telematics will never replace investing in transport infrastructure. It can provide potential improvements to the safety and comfort, but will not increase the efficiency of a road network.

Telematic systems are an integral part of a transport infrastructure. The current rapid development in telematic technologies often comes from private company initiatives.

One should, therefore, remember to build a strategic public-private partnership (PPP), which will allow for implementing effective transport telematics systems in chosen important areas. The safety of transport telematics should be enhanced. A growing scope of applications for telematics creates the need for protection against equally growing criminal threats. A basic telematic architecture, which would be sufficiently protected against external interference, is advisable.

References

1. Mikulski, J.: Charakterystyka ogólna telematycznych systemów transportowych. In: Telematyka Transportu Drogowego, Wydawnictwo ITS, Warszawa, pp. 73–103 (2009)
2. Mikulski, J.: Wizja Rozwoju Inteligentnych Systemów Transportowych w Polsce. Magazyn Autostrady 10, 34–40 (2009)
3. Bartczak, K.: Scenarious for ITS services deployment in Poland. In: Zeszysty Naukowe Politechniki Śląskiej, Gliwice, vol. 1657 (2002)
4. ITS Planning Handbook, ITS City Pioneers Consortium (1998)
5. Miles, J., Chen, K. (eds.): ITS Handbook. PIARC (2004)

Development Trends in Intelligent Transport Systems in Respect to Environmental Protection

Jolanta Joszczuk–Januszewska

Gdynia Maritime University, al. Jana Pawla II 3
81-345 Gdynia, Poland
jolajj@am.gdynia.pl

Abstract. The potential systemic impact of Information and Communication Technologies (ICT) is particularly apparent in the transport sector. ICT can be applied to transport through the development of Intelligent Transport Systems (ITS), in particular. Now, mobilising ICT to facilitate the transition to an energy-efficient, low-carbon economy in ITS is important challenge. For example, ICT can enable energy efficiency improvements by reducing the amount of energy required to deliver a given services by delivering innovative technologies, e.g. emerging solutions in computing such as virtualization, grid computing and cloud computing technologies which are described in this paper.

Keywords: Intelligent Transport Systems, Information and Communication Technologies, virtualization, grid computing, cloud computing.

1 Introduction

Sustainable transport (or green transport) initiative is connected with important challenge. Being able to move about flexibly, safely and affordability is a basic human need. However, current modes of transport are producing about a third of the environmental impacts from human society, especially when powered by fossil fuels. There is a pressing need for multi-model- and intelligent system of transport that will facilitate the choice of low carbon transport options that simultaneously reduce land consumption, congestion, travel times and the health impacts of air pollutants released during travel. This requires both careful planning in the design of communities and the road, rail, trail and public transit that support it, and technological innovations that allow low impact modes of transport to meet human needs effectively. One of specific services are Intelligent Transport Systems (ITS) to provide communication between travellers and the transport network so congestion is managed [1], [2], [3] and [4].

In December 2008, the European Union reiterated its commitment to meeting its energy savings and carbon emissions targets by 2020. These targets only seem achievable if a reduction in energy consumption supported by energy efficient technologies take place. In principle, many innovative technologies are strongly linked with Information and Communication Technologies (ICT)[5].

ICT can have a systemic impact on other sectors of the economy and of society, and can help in providing a basis for sustainable development. The potential systemic

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 183–193, 2010.

impact of ICT is particularly apparent in the transport sector. ICT can be applied to transport through the development of ITS, in particular.

Now, mobilising ICT to facilitate the transition to an energy-efficient, low-carbon economy in ITS is important challenge. For example, ICT can enable energy efficiency improvements by reducing the amount of energy required to deliver a given service by delivering innovative technologies, e.g. emerging solutions in computing such as virtualization, grid computing and cloud computing technologies which are described in this paper.

2 Environmental Protection

Environmental protection is a practice of protecting the environment, on individual, organizational or governmental level, for the benefit of the natural environment and (or) humans. Activism by the environmental movement has created awareness of the various environmental issues. Protection of the environment is needed from various human activities.

For example, the twentieth annual meeting on science and technology for environmental protection of the Society of Environmental Toxicology and Chemistry (SETAC) Europe was in Seville, Spain from 23 to 27 May 2010. The SETAC Europe annual meeting is Europe's biggest meeting on environmental toxicology and chemistry with more than 1500 presentations in parallel platform sessions and poster sessions, participants and scientific speakers from academia, business and government and a blend of scientists and practitioners, researchers and regulators all in attendance.

The programme included eight parallel sessions covering major topics in environmental sciences. The scientific programme included such topics as climate change and marine environment, while the parallel programme is composed of business events, student activities and an annual general assembly.

SETAC is a not-profit, global professional society established in 1979 to provide a forum for individuals and institutions engaged in education, research and development, ecological risk assessment and life-cycle assessment, chemical manufacture and distribution, management and regulation of natural resources, and the study, analysis, and solution of environmental problems. SETAC is an open and democratic organization that operates in a broad social context, reflecting the needs of the environment and people. Application of sound science plays a key role in this process. Membership worldwide comprises about 5,000 professionals in the field of chemistry, toxicology, biology, and ecology: atmospheric, health, and earth sciences; and environmental engineering.

SETAC Europe is one of the four Geographic Units (GU) of the global SETAC, established to promote and undertake activities of SETAC in Europe, and to support activities of SETAC in the Middle East and Africa, currently organized as a Regional Branch to SETAC Europe [6].

SETAC Europe is dedicated to the use of multidisciplinary approaches to examine the impacts of stressors, chemicals, and technology on the environment. The Society also provides an open forum for scientists and institutions engaged in the study of environmental problems, management and regulation of natural resources, education, research and development, and manufacturing. SETAC Europe is incorporated in

Belgium as a not-for-profit organization. The society is governed according its articles of association and by-laws. SETAC Europe maintains its administrative office in Brussels, Belgium.

The primary goals of SETAC Europe are as follows:

- to support the development of principles and practices for protection, enhancement and management of sustainable environmental quality and ecosystem integrity;
- to encourage interactions among environmental scientists and disseminate information on environmental toxicology and chemistry;
- to provide a forum for communication among professionals in government, business, academia, and other segments of the environmental science community and for the protection and welfare of the general public.

3 Three Solutions in Computing as Innovative Technologies in ICT for Development of ITS

Three solutions in computing as innovative technologies in ICT which can be applied to transport through the development of ITS are virtualization, grid computing, and cloud computing.

3.1 Virtualization Technology

Virtualization is the creation of a virtual (rather than actual) version of resource, such as a server, a storage device or network resources.

There are two main areas of information technology where virtualization is making headroads, storage virtualization and server virtualization [7]:

- storage virtualization is the pooling of physical storage from multiple network storage devices into what appears to be a single storage device that is managed from a central console. Storage virtualization is commonly used in storage area networks (SANs);
- server virtualization is the masking of server resources, including the number and identity of individual physical servers, processors, and operating systems from server users. The intention is to spare the user from having to understand and manage complicated details of server resources while increasing resource sharing and utilization and maintaining the capacity to expand later.

3.2 Grid Computing Technology

Grid computing is the act of sharing tasks over multiple computers. Tasks can range from data storage to complex calculations and can be spread over large geographical distances. In some cases, computers within a grid are used normally and only act as part of the grid when they are not in use. These grids scavenge unused cycles on any computer that they can access, to complete given projects [8], [9], [10] and [11].

Grid computing is a form of networking. Unlike conventional networks that focus on communication among devices, grid computing harnesses unused processing cycles of all computers in a network for solving problems too intensive for any stand-alone machine.

In the late 1990s and early 2000s, the idea of grid computing, a type of distributed computing that harnesses the power of many computers to handle large computational tasks, was all the rage, at least among organizations with high-performance computing (HPC) leads.

A well-known grid computing project is the SETI (Search for ExtraTerrestrial Intelligence) @Home project, in which PC users worldwide donate unused processor cycles to help the search for signs of extraterrestrial life by analyzing signals coming from outer space. The project relies on individual users to volunteer to allow the project to harness the unused processing power of the user's computer. This method saves the project both money and resources.

Grid computing does require special software that is unique to the computing project for which the grid is being used.

Grid Computing and the Future of Cloud Computing. Now enterprises are embracing a similar technology – cloud computing, which is a type of computing, comparable to grid computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. Cloud computing and services such as Amazon's Simple Storage Service (S3) provide companies with scalable, high-speed data storage and services at an attractive price [9], [12].

To do this, cloud computing networks large groups of servers, usually those with low-cost consumer PC technology, with specialized connections to spread data-processing chores across them. This shared IT infrastructure contains large pools of systems that are linked together. Often, virtualization techniques are used to maximize the power of cloud computing.

The standards for connecting the computer systems and the software needed to make cloud computing work are not fully defined at present time, leaving many companies to define their own cloud computing technologies. Systems offered by companies, like IBM's "Blue Cloud" technologies for example, are based on open standards and open source software which link together computers that are used to deliver Web 2.0 capabilities.

Cloud computing has started to obtain mass appeal in corporate data centers as it enables the data center to operate like the Internet work through the process of enabling computing resources to be accessed and shared as virtual resources in a secure and scalable manner.

Can cloud computing succeed where grid failed and find widespread acceptance in enterprise data centers? And is there still room for grid computing in the brave new world of cloud computing?

Differences Between Clouds and Grids. While there are many similarities between grid and cloud computing, it is the differences that matter most. Grid computing is better suited for organizations with large amounts of data being requested by a small number of users (or few but large allocation requests), whereas cloud computing is better suited to environments where there are a large number of users requesting small amounts of data (or many but small allocation requests) [12].

"Grids are well suited for complex scientific work in virtual organizations," explained Wolfgang Gentzsch, who was behind Sun's grid efforts and now sits on the board of directors of the Open Grid Forum and is an advisor to the European Union DEISA project. Clouds, on the other hand, are well suited for simple work such as many short-running jobs, he said.

Another key difference between the two: Grids require batch job scheduling or sophisticated policies for allocating jobs, while clouds do not. Also, by their nature, clouds do not require as large an upfront investment, as the cloud provider is responsible for running and maintaining servers.

Gentzsch ascertained that clouds will not replace grids, as grids have not replaced capability HPC, over the last 10 years, as some have predicted and all three technologies have their place. According to Gentzsch we can say that what we will see over the next couple of years is that these different computing nodes will more and more grow together with the World Wide Web and the Internet, until all these resources become one global infrastructure for information, knowledge, computation and communication, the World Wide Grid.

But Kate Keahey offered a slightly different prediction – it is more likely that grids will be re-branded or merge into cloud computing, and grid computing helped create a certain technology reality which made clouds possible. According to her opinion when it comes to IaaS [infrastructure as a service], in five years, since 2010 year, 80 to 90 percent of the computation will be cloud-based. Kate Keahey is a scientist in the Mathematics and Computer Science Division at Argonne National Laboratory who frequently writes about grid and cloud computing. Argonne was the birthplace of grid computing and the Globus project, the de facto grid computing standard.

4 Direct Environmental Impact of ICT for Development of ITS

Carbon footprint (CF) – also named Carbon profile – is the overall amount of carbon dioxide (CO2) and other greenhouse gas (GHG) emissions (e.g. methane, laughing gas, etc.) associated with a product. ICT can be used to increase energy efficiency and reduce carbon dioxide emissions in ITS. Direct energy use of ICT is connected with data centers and equipment in the office [13], [14], [15] and [16]. A data center is a dedicated facility used to house ICT equipment such as servers (computers that offer services over a network) and data storage. Potential areas for reducing the levels of carbon dioxide in data centers is presented in the Table 1.

Data centers are where ICT equipment is most concentrated, but around half of all ICT's energy consumption comes from office equipment including PCs, laptops, printers. This energy use is more difficult to tackle than that of data centers because it relies on the behaviour of many individuals, but some technological improvements are possible (Table 2).

5 New ICT Standard for a Better Environment

To stabilize and eventually reduce CF is an enormous challenge that will require many new technologies, behaviours and practices. New standards will ensure that these technological systems work together well, and provide end-users with the information they need to make informed decisions.

Table 1. Areas for ICT carbon reduction in data centers [17], [18]

Actions	**Rationale**
Server Optimization • Implement storage virtualization and capacity management • Convert existing physical servers to "virtual servers" – partition servers that run in parallel on the same hardware without any interference • Turn off servers outside their service level agreement, subject to a phase loading and chiller unit risk assessment When designing and provisioning new services, create "virtual servers" instead of procuring physical new servers. • Implement a multi tiered storage solution, much of the data spinning on disks today is seldom accessed	Assists in identifying unused servers and disks. Air-conditioning/cooling equipment typically requires at least the same power as the servers they cool, so reducing servers may save twice the power required to run them. Industry practice has been to run a server using only 20% of its capacity. A server which is switched on but idle still requires 50-70% of the power it uses when it is running under maximum load, therefore a single server running at 80% load uses considerably less energy than 4 servers each running at 20% load. Configure several "virtual" servers onto a single server to increase capacity used. Using a single device in this way not only reduces the hardware and support costs but also decreases the energy requirement In traditional data centers, the servers are used inefficiently. For example, a server will still draw around 70-90% of its maximum power usage even when doing no useful work. A technologies called virtualization, grid computing, cloud computing, allow applications that would otherwise be run on several different servers to share one, and in the longer term applications will be rewritten to run naturally using shared resources. This means that servers can run closer to their maximum capacity, which is more efficient. Redundant servers can then be removed.
Identify servers and data disks in the data centre that are running but not providing any services and decommission	A server which is switched on but idle still uses 50-70% of the power used when running at maximum load.
Specify low-power consumption servers with high-efficiency Power Supply Units (80% conversion or better)	Do not over specify system requirements. The higher the specification the more mains power is drawn. Power Supply Units convert mains AC power to the DC power needed by computers. More efficient units minimize the loss of energy from this conversion in the form of heat.

Table 1. (*Continued*)

Consider Blade technology	Research shows that blade servers, networking, storage etc. may require less power than previous generations of rack servers and may require less heat so require less cooling
Data centre audit	Identifies mismatches between the current physical layout and the layout that would maximize the effectiveness of cooling from air conditioning units. Up to a 20% reduction in cooling could be achieved.
Reduce cooling in the data centre to appropriate levels and increase the ambient room temperature	Research has shown that increasing temperatures in data center does not lead to a higher failure rate as was previously thought. Over 50% of the power associated with the data centre is used for cooling the ICT equipment.
Reduce and re-use of equipment that is no longer required but is still serviceable	Energy is required to manufacture, distribute and recycle equipment as well as to use it. Extending its use or seeking its re-use elsewhere will save energy as well as purchase and disposal costs.
Ensure that equipment can be serviced (e.g. disks, RAM)	Servicing and assessing equipment rather than replacing it could minimize energy consumption through manufacture and disposal
Remove un-necessary or duplicated data/information	The more data stored the more equipment and therefore power is required

In October 2009, it was announced by the British Standards Institute (BSI) that a new standard was being developed [19]. The new PAS 2060, built upon existing ISO 14000 and PAS 2050 standards, is aimed at providing a consistent approach to assess carbon-neutrality claims.

Carbon neutral means that – through a transparent process of calculating emissions, reducing those emissions and offsetting residual emissions – net carbon emissions equal zero.

The BIS not only believes that PAS 2060 will reduce greenhouse gas emissions but also to encourage broader actions toward climate control and carbon reduction management.

Table 2. Improvements in the office for ICT carbon reduction [13]

Improvements	Rationale
More efficient components	The micro-processors within electronic equipment require energy both to operate and for cooling fans. Advances in chip design (such as 'multi-core' processors) can save 30-60% of the energy used by the processor if software is written to take advantage of this capacity.
Power management	Almost all computers now have a low power mode which they can enter automatically after a period of user inactivity. In such modes they will consume very little power but can often be woken up within seconds. Power management options are sometimes enabled as a default.
Laptop computers	These can sometimes use as little as a third of the energy of a desktop. Some organizations are replacing desktops in the office with laptops. Similarly, flat screen monitors are much more efficient than old-fashioned cathode ray tubes.
Thin client technology	A Thin client is less complex than a PC and contains fewer components, increasing its life over that of a normal PC and reducing maintenance and support costs and thus energy consumption. However additional energy is required to support the greater bandwidth necessary for connection to its server as well as to run the server and its supporting air-conditioning equipment. Implementation of thin client technology should be balanced against potential increases in server usage.
Multi-functional devices	Printers now often include scanning, copying and fax functions, which is more efficient than running several separate devices. The number of printers per person can also be reduced, and printers can be set to enforce double-sided printing to save paper and use of the printer.

6 European Union Steers Its Way towards Interoperability of Intelligent Transport Systems

Maximizing the potential of ITS for reducing emissions will significantly depend on interoperability of ITS systems.

A coherent EU–wide framework for interoperable transport systems is the aim of a proposal adopted by EU transport ministers today. Following negotiations with the European Parliament, the European Council today approved a proposal for a directive on the deployment of ITS in the field of road transport [20].

The Council adopted its position at first reading on the ITS directive proposal. The text will now be transmitted to the Parliament for a second reading, which is due to take place in June. If the Parliament accepts the Council's position, the directive will be considered adopted in this form.

The objective of the draft directive is to accelerate and coordinate deployment of interoperable ITS in road transport, including interfaces with other transport modes, by creating the necessary conditions and mechanisms through a coherent EU–wide framework.

The Council agreed on a two-step approach paving the way for swift introduction of ITS through EU legislation: first, the Commission adopts the necessary specifications for the ITS applications and services concerned; then, within 12 months and, where appropriate, after an impact assessment, the Commission presents a proposal for deployment of those ITS to the European Parliament and the Council, which will jointly decide for or against the proposal. In any case, EU Member States will have the final say on whether or not to deploy an ITS application or service on their territory. However, if they do so, they must respect the specifications adopted under the directive.

Intelligent transport systems integrate innovative ICT with transport engineering and traffic management. They are intended to enhance road safety, improve transport efficiency and reduce the environmental impact of transport. To foster the deployment of ITS, the directive defines priority areas and priority actions within those areas. The Commission will have the task of adopting specifications for the actions planned in the priority areas.

The priority areas and corresponding main actions outlined in the draft directive include:

- optimal use of road, traffic and travel data priority actions: EU–wide multimodal travel and real–time traffic information services as well as road safety related minimum universal traffic information services;
- continuity of traffic and freight management ITS services (actions: e.g. ensuring information flow; tracking and tracing of freight);
- ITS road safety and security applications priority actions: EU–wide eCall services as well as reservation and information systems services for safe and secure parking places for trucks and commercial vehicles;
- linking the vehicle with the transport infrastructure (actions: e.g. systems for exchange of data or information between vehicles, infrastructures and between vehicle and infrastructure).

The propositions mentioned above are the subjects of the Directive of the European Parliament and of the Council on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport [21].

7 Conclusions

- Sustainable transport (or green transport) refers to any means of transport with low impact on the environment. One of specific services are Intelligent Transport Systems.

- By delivering innovative technologies, ICT can reduce wasteful consumption of energy. Emerging solution in computing such as virtualization, grid computing, and cloud computing technologies promise to reduce redundancies existing in today's (May 2010) systems.
- Direct environmental impact of ICT for development of ITS is connected with data centers and equipment in the office. Areas of ICT carbon reduction in data centers and improvements in the office are the new solutions for the environmental protection.
- PAS 2060 is the new ICT standard for a better environment.
- Maximizing the potential of ITS for reducing emissions will depend on interoperability of ITS. For this reason coherent EU–wide framework for interoperable transport systems is the aim of a proposal adopted by EU transport ministers nowadays.

References

1. McDonald, M., et al.: Intelligent Transport Systems in Europe. World Scientific, Singapore (2006)
2. Standardization Activities for Intelligent Transport Systems. ITU-T Technology Watch Report 8 (2008)
3. Weldon, G., Griffiths, M.: Intelligent Transport Systems. Postnote Number 322. The Parliamentary Office of Science and Technology, London (2009)
4. Williams, B.: Intelligent Transport Systems Standards. Artech House Publisher, Norwood (2008)
5. Commission of the European Communities. Communication from the commission to the European Parliament, COM 111, Brussels (2009)
6. `http://www.seville.setac.eu/general_info`
7. `http://searchservervirtualization.techtarget.com`
8. Bekakos, M.P., et al.: Grid Technologies. Emerging from Distributed Architectures to Virtual Organisations. WIT Press (2006)
9. Distributed Computing: Utilities, Grids & Clouds. ITU-T Technology Watch Report 9 (2009)
10. Pearce, S., Nath, C.H.: Grids and e-Science, Postnote Number 286. The Parliamentary Office of Science and Technology, London (2007)
11. Wu, Z., Chen, H.: Semantic Grid: Model, Methodology, and Applications. Springer, Heidelberg (2008)
12. `http://www.enterprisestorageforum.com//article.php/3859956/Grid-Computing-and-the-Future-of-Cloud-Computing.htm`
13. Griffiths, M.: ICT and CO2 Emissions, Postnote Number 319. The Parliamentary Office of Science and Technology, London (2008)
14. ICTs and Climate Change, ITU-T Technology Watch Report 3 (2007)
15. ICTs for the Environment, ICT World Today Spring (2009)
16. Information and Communication Technology. Key Technology Area. 2008-2011, The Technology Strategy Board UK (2008)
17. CTO COUNCIL Green ICT Working Group. Areas for Potential Carbon Reduction. Extended List of Action 28/10/2008

18. Going Green, I.T.: Initiatives by Five Economics, ICT World Today Spring (2009)
19. http://www.talkstandards.com/recent-trends-in-green-ict
20. http://www.eubusiness.com/news-eu/transport-system-directive.105
21. Directive of the European Parliament and of the Council, Council of the European Union, 6103/10, Brussels (May 4, 2010)
22. Mikulski, J., Kwaśny, A.: Role of telematics in reducing the negative environmental impact of transport. In: Mikulski, J. (ed.) Communications in Computer and Information Science: Transport Systems Telematics, vol. 104 (in print, 2010)

Evolutionary Algorithms Find Routes in Public Transport Network with Optimal Time of Realization

Anna Piwonska and Jolanta Koszelew

Technical University of Bialystok
Computer Science Faculty
Wiejska 45A, 15-351 Bialystok, Poland
{a.piwonska,j.koszelew}@pb.edu.pl

Abstract. This paper presents a new evolutionary algorithm, called Routes Generation Evolutionary Algorithm with Knowledge (RGEAwK), for determining routes with optimal travel time in graph which models the public transport network. The method was implemented and tested on the real transport network. The effectiveness of the method was compared with Routes Generation Matrix Algorithm (RGMA) [8]. This comparison is based on the experimental results which were performed on realistic data. The paper ends with some remarks about future work on improving of RGEAwK.

Keywords: Public transport network, time-dependent graph, optimal routes, evolutionary algorithm.

1 Introduction

There are many commercial computer systems which solve optimization problems in transport networks: BUSSMAN, HOT or HASTUS [1], [2]. However, it is very difficult to find optimal routes in public transport systems which take into consideration users' preferences. Effective algorithms which generate optimal routes in a public transport network are the heart of journey planners [4], [3]. The users of such system determine the starting point and the destination point of travel, start time, their preferences and as a result, system returns information about optimal routes. In practice, public transport users' preferences may be various, but the most important of them are: a minimal travel time and a minimal number of transfers. Because of the fact that graph which models the public transport network has time-dependent dynamic weights, the problem of finding routes with a minimal travel time is called *NPc* [5], [6].

Authors present a new, original evolutionary algorithm, called RGEAwK for determining routes with optimal travel time. The method was implemented and tested on the real public transport network. The effectiveness of the method was compared with RGMA algorithm [7], [8]. RGMA realizes the label-setting strategy and uses two special matrices, called transfer matrix and minimal distance matrix and exploits the following assumption: the routes with minimal (or close to minimal) number of transfers or with minimal length (number of bus stops on the route) have probably the optimal time of realization.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 194–201, 2010.

The previous version of RGEAwK was RGEA algorithm [9]. The main problem with RGEA was insufficient performance for routes with very long distances. It was the motivation for the authors to try to improve the RGEA. The main improvement was incorporating the knowledge about transport network to the algorithm. The improved version, RGEAwK, is described in this paper.

Next section includes definition of optimal routes generation problem and the description of public transport network model. RGEAwK is described in section 3. In section 4 authors present the comparison of effectiveness of two methods: RGMA and RGEAwK in two aspects: time of realization of routes (taking into account the number of transfers in routes) and time complexity. This comparison is based on experimental results which were performed on realistic data for Bialystok city. The paper ends with some remarks about future work on improving of RGEAwK.

2 Network Model and Problem Definition

A transportation network in our model is represented as a bimodal weighted graph $G = \langle V, E, t \rangle$ [10], where V is a set of nodes, E is a set of edges and t is a function of the weights. Each node in G corresponds to a certain transport station. We assume, for simplification, that there is only one kind of public transportation - the bus, so each node corresponds to a bus stop. This assumption does not limit the applications of the presented methods. We also assume that bus stops are represented with numbers from 1 to n. The directed edge $(i, j) \in E$ is an element of the set E, the bus line number l connects the stop number i as a source point and the stop number j as a destination. One directed edge called bus link corresponds to one possibility of the connection between two stops. Each edge has a weight t_{ij} which is equal to the travel time (in minutes) between nodes i and j which can be determined on the base of timetables (Tab. 1). A set of edges is bimodal because it includes, besides directed links, undirected walk links. The undirected edge $\{i, j\} \in E$ is an element of the set E, if walk time in minutes between i and j stops is not greater then $limit_w$ parameter. The value of $limit_w$ parameter has a big influence on the number of network links (density of graph). The t_{ij} value for undirected edge (i, j) called walk link is equal to walk time in minutes between i and j stops. We assume, for simplification, that a walk time is determined as an Euclidian distance between bus stops. It is very important to take the walk links into consideration, because public transportation network is very rare in small cities or peripheral districts of big cities, and walk links increase chance of finding at least one route between bus stops which are located in such regions.

A graph representation of public transportation network is shown in Fig. 1. It is a very simple example of the network which includes only nine bus stops. In the real world the number of nodes is equal to 1500 for the city with 250 thousands of inhabitants.

Formal definition of our problem is the following. At the input we have: the graph of transportation network, timetables - times of departures for each stops and each line, travel starting point (s), destination point of the travel (d), starting time of the travel ($time_s$), number of the resulting paths (k), maximum number of transfers (max_t) and limit for walk links ($limit_w$). The max_t parameter is required only for RGMA method. At the output we want to have the set of resulting routes, containing at most k quasi-optimal paths with minimal time of realization (in minutes), with at most max_t transfers - assumption important only for RGMA.

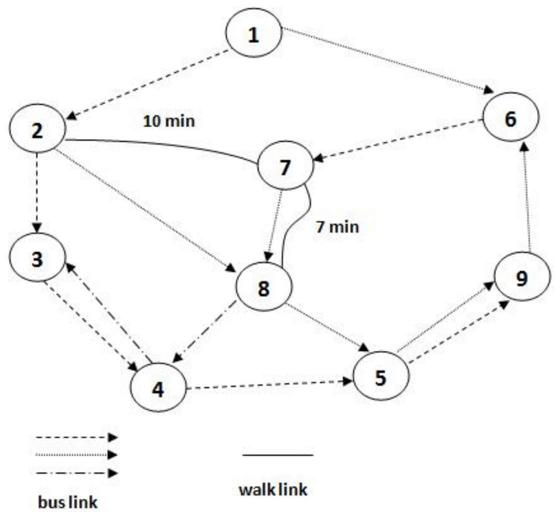


Fig. 1. Representation of a simple transportation network (different styles of lines mark different bus links; solid lines mark walk links)

Table 1. Timetables for each bus stop presented in Fig.1

nr	Line	Times of departure		
1	Dash	5:55	6:10	
	Dash-dot	6:01	6:13	6:25
2	Dash	5:58	6:13	
	Dot	6:05	6:20	8:30
	Solid	10 min		
3	Dash	6:03	6:18	
	Dash-dot	6:09	6:21	6:33
4	Dash	6:05	6:29	
	Dash-dot	6:07	6:19	6:31
5	Dash	6:08	6:23	
	Dot	6:11	6:26	6:36
6	Dash-dot	6:03	6:15	6:27
	Dot	6:15	6:30	6:40
7	Dash-dot	6:04	6:16	6:27
	Solid	10 min		
	Solid	10 min		
8	Dash-dot	6:05	6:17	6:29
	Dot	6:08	6:23	6:33
	Solid	10 min		
9	Dash	6:10	6:25	
	Dot	6:13	6:28	6:38

The t_{ij} values are marked in Fig. 1 only for walk links, because only this kind of links is time independent. Weights for bus links are strongly dependent on the starting time parameter and can be computed only during the realization of algorithm. We can determine t_{ij} values for bus links using timetables. We assume, for simplification, that each bus line has the same timetables for each weekday.

3 RGGAwK for Discovering Routes

The pseudocode of RGEAwK is presented below.

```
  procedure RGEAwK(m,s,d,max_t,time_s,limit_w,P,ng,pr,k)
    begin
      generate m routes from s to d with minimal number
      of tranfers (not greater than max_t) starting at
      time time_s, with maximal walk links equal to
      limit_w;
      include m routes to an initial population of
      RGEAwK;
      generate P-m routes from s to d in a random way;
      compute fitness function F for each individual;
      for i:=1 to ng do
      begin
       choose with probability pr a genetic operator:
       crossover or mutation;
       if operator=crossover
       begin
         choose two individuals according to the F value;
         cross parent individuals, if possible;
         compute fitness function F for new individuals;
         add offspring individuals to the population;
       end
      else
      begin
        randomly choose one individual;
        mutate the individual;
        compute fitness function F for mutated
        individual;
        add mutated individual to the population;
     end;
  end;
choose k best routes from the final population;
end;
```

The RGEAwK starts with a population of P solutions of a given problem. The initial population is generated in a special way: m individuals are computed as routes from bus stop s to bus stop d, starting at time $time_s$, with minimal number of transfers not greater than max_t and with maximal walk links equal to $limit_w$ [10]. Determining these routes is not a difficult task, since they are computing on a graph without time-dependent link (a network without timetables represented by a graph without weights). The rest of the population is generated in a random way. The next step is to evaluate individuals in the initial population by means of the fitness function F. The fitness function should estimate the quality of individuals, according to the time of realization of the tour. After fitness evaluation, the RGEAwK starts to improve initial population through ng applications of crossover and mutation.

In every generation we first choose with probability *pr* between crossover and mutation. In the case of crossover, we first select two parent individuals, according to the fitness value: the better an individual is, the bigger chance it has to be chosen. Since chromosomes lengths are different, we presented a new heuristic crossover operator, adjusted to our problem [9]. In the first step we test if crossover can take place. If two parents do not have at least one common bus stop, crossover cannot be done and parents remain unchanged. Crossover is implemented in the following way. First, we choose one common bus stop, it will be the crossing point. Then we exchange fragments of tours from the crossing point to the end bus stop in two parent individuals. After crossover, we must correct offspring individuals in two ways. First we eliminate so called ”bus stop loops”, then we eliminate so called ”line loops” [9]. The next step is to compute fitness function for these new individuals. Finally, offspring individuals are added to the population, they do not replace their parents.

In case of mutation, we first choose randomly one chromosome. The next step is to randomly select two bus stops, denoted as $s1$ and $d1$ from the route ($s1,d1 \neq s,d$). Then we generate m routes from $s1$ to $d1$ with minimal number of transfers not greater than max_t and with maximal walk links equal to $limit_w$ [10]. From these m routes we select a route with minimal time of realization. This best route exchanges the fragment of a route from $s1$ to $d1$ in a chromosome being mutated. Then we compute fitness function for this individual and add it to the population.

4 Experimental Results

There was a number of computer tests conducted on real data of transportation network in Bialystok city. This network consists of about 700 bus stops, connected by about 30 bus lines. There was an assumption about length of walk links $limit_w$ - they were limited to 15 minutes. The value of $limit_w$ is very important because it influences the density of network. The bigger value of $limit_w$, the more possibilities of walk links in a network. The density of network is of a key importance for time complexity of algorithms.

The parameters of RGEAwK were: $P = 200$, $m = 2$, $max_t = 5$, $pr = 0.5$, $k = 5$ and $ng = 500$. We performed three kinds of tests. We examined routes from the centre of the city to the periphery of the city (set C - P), routes from the periphery of the city to the centre of the city (set P - C) and routes from the periphery of the city to the periphery of the city (set P - P). Each of these sets includes routes which are difficult cases for both algorithms. The first issue is the lenth of the distance from s to d, the second is a rare density of the network in s or d localization. Rare density of network could be the reason of high value of approximation error for routes.

The examples of routes from each set, generated by RGEAwK, are presented in Tab. 2, 3, 4. The last four columns in each table denote time (-t) and number of transfers (-tr) of the best route generated by RGMA and RGEAwK, respectively.

In case of routes of C - P type, that RGEAwK has less average time of realization than RGMA in 70% of examined routes. The average difference is equal to 12 minutes. Additionally, in 53% cases where RGEAwK was not worse than RGMA, the number of transfers in RGEAwK was only one more than in RGGA, maximally. RGEAwK has never been worse than RGMA according to time of realization.

We can observe on the base of tested routes that RGEAwK has less average time of realization than RGMA in 40% of examined routes of *P* - *C* type. The average difference is equal to 21 minutes. Additionally, in 40% cases where RGEAwK was not worse than RGMA, the number of transfers in RGEAwK was only two more than in RGGA, maximally. RGEAwK has never been worse than RGMA according to time of realization.

In case of routes of *P* - *P* type, that RGEAwK has less average time of realization than RGMA in 40% of examined routes. The average difference is equal to 18 minutes. Additionally, in 30% cases where RGEAwK was not worse than RGMA, the number of transfers in RGEAwK was three more (only in one case) than in RGGA, maximally. RGEAwK has never been worse than RGMA according to time of realization.

Table 2. The examples of routes from set *C* - *P*; $time_s$ = 15:30

nr	s	d	RGMA-t	RGGAwK-t	RGMA-tr	RGGAwK-tr
1	Malmeda	Dojlidy G.	00:28	00:28	1	1
2	Akademia Med.	Nowodworce	01:42	01:28	1	3
3	Bazantarnia	Wiadukt	01:10	00:46	1	2
4	Berlinga	Dojlidy G.	00:58	00:58	1	1
5	Botaniczna	Chelmons.	00:45	00:45	1	1
6	Branickiego	Dojlidy G.	01:25	00:28	0	1
7	Hetmanska	Dojlidy G.	01:25	00:58	1	2
8	Hotel Gromada	Mysliwska	01:24	00:48	0	1
9	K.E.N./Kollataja	Nowodworce	01:42	01:42	2	2
10	Kino Pokój	Silikaty	00:31	00:31	1	1
11	Klub Rozrywki	Klepacze	00:53	00:53	1	1
12	Koscielna	Silikaty	00:31	00:31	1	1
13	Kosciol s. Woj.	Makro	00:49	00:49	1	1
14	Monte Cassino	Kleosin	00:54	00:54	1	1
15	Wierzbowa	Silikaty	01:46	01:16	1	2
	Nr of winners		9	15	9	15

Table 3. The examples of routes from set *P* - *C*; $time_s$ = 7:30

nr	s	d	RGMA-t	RGGAwK-t	RGMA-tr	RGGAwK-tr
1	Nowodworce	Malmeda	03:08	01:35	2	3
2	Zagorki	Wierzbowa	02:17	01:42	2	3
3	Silikaty	Akademia M.	01:31	00:51	1	2
4	Ksiezyno	Czestochow.	01:16	00:52	1	2
5	Wiadukt	Fabryczna	00:51	00:38	1	2
6	Ksiezyno	Gajowa	01:07	01:07	1	1
7	Baranowicka	Grottgera	00:35	00:35	2	2
8	Ksiezyno	Hala Jagiell.	00:49	00:49	1	1
9	Nowodworce	Hetmanska	01:37	01:32	2	1
10	Dojlidy G.	K.E.N./Kollataja	01:17	00:55	1	2
11	Produkcyjna	Kalinowskiego	00:26	00:26	1	1
12	Niemenska	Kino Pokoj	01:28	01:28	2	2
13	Dojlidy G.	Klub Rozrywki	01:44	01:03	1	2
14	Zagorki	Koleowa PKP	01:21	00:39	1	2
15	Dojlidy G.	Kopernika	00:36	00:36	1	0
	Nr of winners		5	15	5	15

Table 4. The examples of routes from set *P* - *P*; $time_s$ = 15:30

nr	s	d	RGMA-t	RGGAwK-t	RGMA-tr	RGGAwK-tr
1	Wiadukt	Silikaty	00:49	00:49	1	1
2	Klepacze	Dziesięciny	00:49	00:49	1	1
3	Nowodworce	Zielone Wzgorza	01:50	01:50	1	2
4	Klepacze	Ksiezyno	02:00	00:58	1	4
5	Gajowa	Chelmonsk.	00:57	00:57	1	1
6	Chelmonsk.	Bialostocka	01:34	01:24	1	2
7	Gajowa	Mysliwska	00:55	00:55	1	1
8	Polmos	Wiadukt	00:55	00:46	1	2
9	Gajowa	Nowodworce	01:42	01:42	1	1
10	Silikaty	Dubois	01:36	01:36	1	1
11	Wiadukt	Fabryka Dyw.	04:28	01:38	1	2
12	Nowodworce	Kleosin	01:17	01:17	2	2
13	Gajowa	Polmos	01:02	01:02	1	1
14	Silikaty	Produkcyjna	01:04	00:56	1	0
15	Polmos	Swobodna	01:09	00:51	1	2
	Nr of winners		9	15	9	15

The last experiment was focused on comparison of time complexity of algorithms. The results are presented in Fig. 2.

In this experiment we tested examples of routes with a minimal number of bus stops, between 5 and 30. The set of tested routes was generated by BFS-method on the base of the structure of the network (graph without weights) [7]. The minimal number of bus stops is marked on *x* axis. On *y* axis we denoted the time of realization of algorithm in miliseconds (processor Pentium 3.0 GHz). Precise analysis showed that RGEAwK is on average 3143 ms better than RGMA.

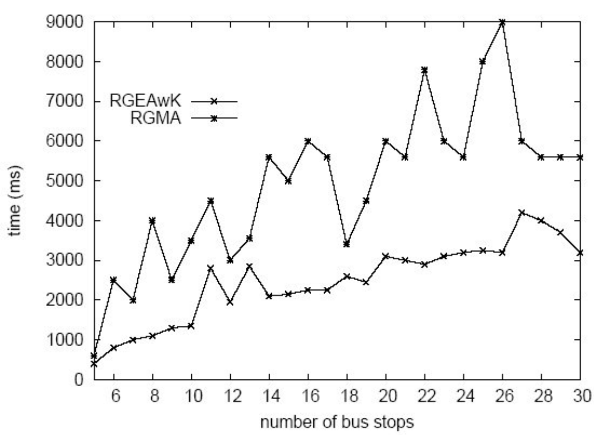


Fig. 2. The comparison of time complexity of RGEAwK and RGMA

5 Conclusions

The reason for developing a new evolutionary algorithm was too high time complexity and insufficient quality of resulting routes generated by RGMA and RGEA [7], [9]. The authors' motivation was to try to improve the RGEA. The main improvement was incorporating knowledge about transport network to the algorithm.

Computer experiments have shown that improved algorithm, RGEAwK, performs much more better than RGMA and significantly faster.

Future work will be concentrated on testing RGEAwK on transport network for big metropolises such as Warsaw or Pol. Gornoslaski Okrag Przemyslowy (Silesian Industrial Region). If tests show poor performance of RGEAwK the new heuristics must be added to the algorithm. The proposal of improvement which can be considered includes to the algorithm information about geographic location of start and destination bus stops and topology of network.

Acknowledgments. This research was supported by S/WI/2/2008 and S/WI/3/2008.

References

1. Rousseau, J.M.: Scheduling Regional Transportation with HASTUS. In: Proceedings of CAPST 2000, Berlin (2000) (electronically published)
2. Serna Delgado, C.R., Bonrostro Pacheco, J.: MINMAX Vihicule Routing Problems: Application to Schol Transport in the Providence of Burgos (Spain). In: Proceedings CASPT 2000. Berlin (2000) (electronically published)
3. Wu, Q., Hartley, J.K.: Accommodating User Preferences in the Optimization of Public Transport Travel. International Journal of Simulation Systems, Science and Technology: Applied Modeling and Simulation 5(3-4), 12–25 (2004)
4. Tulp, T.: CVI: Builder of Dutch Public Transportation System(1993) (unpublished)
5. Hansen, P.: Bicriterion path problems. In: Fandel, G., Gal, T. (eds.) Multiple criteria decision making: theory and applications. Lecture Notes in Economics and Mathematical Systems, vol. 177, pp. 236–245. Springer, Heidelberg (1980)
6. Safer, H.M., Orlin, J.B.: Fast approximation schemes for multicriteria combinatorial optimization. Technical Report No. 3756-95. Cambridge, Massachusetts Institute of Technology, Sloan School of Management (1995)
7. Koszelew, J.: Two methods of quasi-optimal routes generation in public transportation network. In: Proceedings of 7th International Conference on Computer Information Systems and Industrial Management Applications: CISIM 2008, pp. 231–236. IEEE Computer Society, Los Alamitos (2008)
8. Koszelew, J.: Approximation method to route generation in public transportation network. Polish Journal Environment Studies 17(4C), 418–422 (2008)
9. Koszelew, J., Piwonska, A.: A new genetic algorithm for optimal routes generation in public transport network. In: Proceedings of 13th International Conference on System Modelling Control: SMC 2009, Lodz University of Technology (2009)
10. Koszelew, J.: The Theoretical Framework of Optimization of Public Transport Travel. In: Proceedings of 6th International Conference on Computer Information Systems and Industrial Management Applications: CISIM 2007, pp. 65–70. IEEE Computer Society, Los Alamitos (2007)

The Safety Related Software for Railway Control with Respect to Automatic Level Crossing Signaling System

Andrzej Lewiński[1] and Katarzyna Trzaska–Rycaj[2]

[1] Faculty of Transport and Electrical Engineering, Technical University of Radom, Malczewskiego 29 St, PL26600 Radom, Poland
a.lewinski@pl.radom.pl
[2] Technical School of Communication, Ułanów 3 St, PL31450 Cracow, Poland
krycaj@tk.krakow.pl

Abstract. The paper deals with design problems of correct and high reliable software for railway traffic control systems. The correct software (corresponding to formal or semi-formal criteria) has an important part in safety related (SIL4) railway control systems. The paper treats about actual state of art in design of safety related software for railway application. The proposed methods, recommended by CENELEC and UIC are introduced to example of automatic level crossing signaling system.

Keywords: Software safety and reliability, level crossing signalling systems, UML modelling, railway traffic control systems.

1 Introduction

Designing a safety-related software in railway computer systems is vitally important for correct operation of the entire railway system. Performing tasks defined by the programmer is the software's main function e.g. control of the signalling on level crossings or ensuring the safety of railway traffic. In order to ensure safety of rail communication users the software applied in railway computer systems must be reliable.

Proper formulation of requirements for the future railway system is essential. It has been proved that most errors (above 50%) arise already at the stage of specifying requirements of the design and at the engineering stage (above 25%). Errors at the coding stage make less than 10% of all errors arising in the production cycle [3].

Engineering, implementation and exploitation of the software for railway traffic control systems are subject to European PN-EN 50128 standard developed by CENELEC. In control systems which should meet high security requirements, the rule of fail-safe assurance is applied. In railway technology this rule has been used for quite a long time in order to ensure reliability and safety in railway traffic control systems functioning. The rule of fail-safe assurance can be expressed in the following way: "failures of the equipment or software errors in the data processing system can lead to unserviceability that may cause damage to the system's condition, however, the system should always be safe, i.e. such damage should not threaten the controlled process or the environment". This means that a single damage does not cause a dangerous situation and should be detected in the properly short time. The above mentioned rule of

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 202–209, 2010.

safety assurance in the case of damages might be applied in two forms, different as regards safety assurance, i.e. in the form of direct or indirect safety assurance.

The direct safety assurance method implies that all equipment failures causing functional disturbances directly bring the system to a safe condition. The indirect fail-safe safety assurance method is a modification of the method described above which allows to assure reliability by introducing additional supervisory and control functions to the system. This method is implemented by means of a multichannel control logic in which channels are fully reliable and therefore are supervised by a special comparator system, by first failure detector and by a system disabling the damaged channel, which control compatibility of output conditions.

2 Software Engineering Methods and Status of Works Related to Safety Related Software for Railways Systems

The software life cycle represents the recurrent totality of actions undertaken from the disclosure of the need for system construction to the end of its utilization [6]. Among cycles of the software engineering the classical models such as the cascade model, the spiral model, the evolutionary model and the V model are the most common. Unfortunately, as it turned out in practice, classical models are not error-free. Additionally, the situation becomes more complicated due to the users' growing requirements, the greater specialization and the complexity of systems. That is why contemporary models, often connected with object-oriented methodology, have gained importance lately. They include: the model of the object-oriented production cycle or the Checkland's model.

Research that has been done on similar issues concerned mainly determining reliability characteristics in systems of signalling devices used in traffic and modelling of fail-safe systems using e.g. Petri's networks or Markov's chains. Also, several dissertations on software engineering methods for fail-safe systems have been published. [7]. They mainly concerned application and role of methods such as: Petri's networks, Z, SDL, or UML in safe software engineering.

Software engineering methods may be classified in respect of their degree of formalization, i.e. the description's conciseness and accuracy and the way the system's different aspects are presented. When applying the first criterion, the following methods may be discerned: informal, semi-formal and formal.

3 Safety Related Software for Level Crossing Signaling Systems

As far as the role of control is concerned, railway traffic control system is a set of: material objects (e.g. executive devices), abstract objects (e.g. the set of control software), attributes (e.g. conditions or properties of objects) and relations between attributes, describing control functions which ensure safe execution of the assigned tasks[11].

Table 1. Software architecture [2]

Techniques/measure	Safety Integrity Levels SIL	
	3	**4**
Defensive programming	HR	HR
Fault detection&diagnostics	HR	HR
Error Correcting Codes	NR	NR
Error Detecting Codes	HR	HR
Failure Assertion Programming	HR	HR
Safety Bag Techniques	R	R
Diverse Programming	HR	HR
Recovery Block	R	R
Backward Recovery	NR	NR
Forward Recovery	NR	NR
Re-try Fault Recovery Mechanisms	R	R
Memorizing Executed Cases	HR	HR
Artificial Intelligence – Fault Correction	NR	NR
Dynamic Reconfiguration of software	NR	NR
Software Error Effect Analysis	HR	HR
Fault Tree Analysis	HR	HR

Table 2. Software requirements specification [2]

Techniques/measure	Safety Integrity Levels SIL	
	3	**4**
Formal methods including for example: CCS, SP, HOL, OBJ, Temporal Logic, VDM, Z and B.	HR	HR
Semi – formal methods	HR	HR
Structured methodology including for example: JSD, MASKOT, SADT, SSADM, and Yourdon.	HR	HR

Where in both Table 1 and Table 2: HR – high recommended, R – recommended, NR – not recommended.

3.1 Description of the SSP System

In order to describe the automatic level crossing signalling system (SSP) a similar pattern can be used (Fig. 1). Traffic control on level crossings is a process which has an effect on objects of the system, ensuring safe motor and railway traffic on the crossing. Time and processes which change the status of SSP system are the control attributes. The aim of the control within this system is to ensure correct and safe traffic on the level crossing [11]. Thus, in order to perform such tasks all elements mentioned in Fig. 1 have to operate correctly.

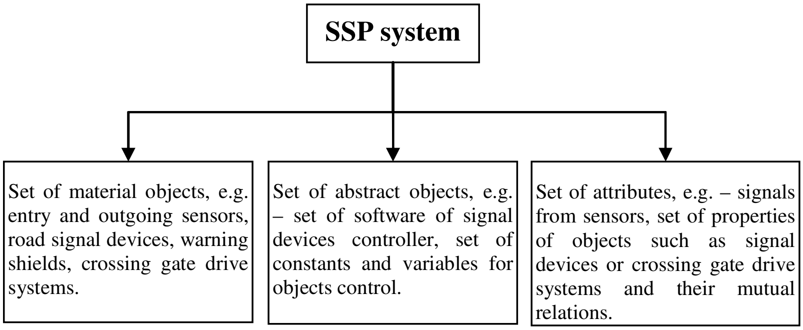


Fig. 1. Description of SSP system. Source: own elaboration based on [11].

Operation of the SSP computer system must be constantly controlled by cyclic- sequential software. The system must always be ready to activate a warning. The algorithm of the control should be performed in at least two synchronized drivers. Software for drivers should be varied and should be designed by different teams of programmers. Moreover, software should contain a wide range of self-testing mechanisms.

To give an example, in the SPA-4 level crossing signalling system produced by Bombardier Transportation, safety of the system software was guaranteed thanks to the use of several different methods, such as:

- diversification of software structures in two independently operating A and B channels. Two independent teams of programmers for drivers in individual channels designed control software. Different RAM areas were used. In addition, different numbers of flags and registers were used.
- SPA-4 system contains the range of self-test mechanisms, enabling formal testing of the software cycle as well as testing of system devices. Testing of the software operation cycle consists in cyclic exercise of procedures checking and verifying the correctness of current values of parameters. Additionally, in each software operation cycle, integrity of the algorithm of control software is analyzed.

3.2 SSP System Trial Implementation with the Use of Formal Methods

For a proper functioning of the system, it is already at the engineering stage that the author should employ formal and semiformal methods recommended by CENELEC (table 2) [2]. Also a method based on the standardized modelling language of UML computer systems may be indicated here. The model of the SSP automatic level crossing system should: enable control assignment modelling of the SSP system, perform mapping of the structure of the SSP system which consists of the control system and objects, and mapping of the control system, i.e. a system controlling the level crossing signalling [11].

The SSP automatic level crossing signalling system is defined as presented below:

$$S_{SSP} = (B, S_S, U_S) \tag{1}$$

where:
B – set of devices (objects) in the system;
S_S – control system managing entire SSP system as well as objects B;
U_S – set of control functions.

The entire set of SSP system devices can be divided into following subsets:

$$B = \{B_{ST}, B_{SK}\} \tag{2}$$

where:
B_{ST} – subset of objects controlled by system, $B_{ST} = \{b_{st1}, b_{st2}, \ldots, b_{stn}\}$
B_{SK} – subset of objects supervised by system, $B_{SK} = \{b_{sk1}, b_{sk2}, \ldots, b_{skm}\}$
where: n and m determine accordingly numerousness of subsets of controlled objects (e.g. acoustic and sound signal device for those who drive the road vehicles) and subset of supervised objects (e.g. – railway sensors) within SSP system.

Model MS_S of the control system S_S can be represented in form of:

$$MS_S = \{M_{STA}, M_{DYN}\} \tag{3}$$

where:
M_{STA} – static model of the system;
M_{DYN} – dynamic model of the system.

Static model M_{STA} contains permanent information of the system, such information concern required properties of objects included in the SSP system. Dynamic model M_{DYN} contains information which is subject to changes in time; their status depends on the real situation in the level crossing and can change over time.

To describe the SSP program, a model allowing a binary relation assigned to the instruction to transform the state of the variables can be introduced. These variables will change when running the instruction [5].

The SSP control program is a composition of *m* modules {P1, P2 … Pi … Pm} being expressed as following:

$$P_{SSP} = F(P1, P2, \ldots, Pi, \ldots, Pm) \tag{4}$$

The SSP modules can be either blocks, subprograms, procedures or functions. The formal methods of the program analysis introduce the features defined in a mathematical theory of programming. Each of the SSP modules can be characterized by the I-O specification, as in:

$$\langle \varphi, P_{SSP}, \psi \rangle \tag{5}$$

where:
φ - pre-condition describing the input variables at the initial stage, before applying the P_{SSP} program,
ψ - post-condition describing the expected outcomes that is the values of the variables at the final stage, after applying the P_{SSP} program.

The correctness of $<\varphi, P_{SSP}, \psi>$ is to be proven by applying a proper mathematical model which can be expressed as following (5):

$$C(\langle \varphi, P_{SSP}, \psi \rangle) \tag{6}$$

Calculating the outcome results is difficult for the entire SSP program and for that reason the condition of correctness can be decomposed to a set of local conditions:

$$C(\langle \varphi, P_{SSP}, \psi \rangle) \equiv \{C(\varphi_i, P_{SSPi}, \psi_i)\}_{i=1,2,\ldots,m} \tag{7}$$

Such condition of the SSP correctness can be additionally decomposed to the compound instructions. Quite often in the highway crossing signaling one can find the equivalent programs being used. For the error detection the results of *m* programs running simultaneously are being compared.

3.3 Application of UML Language for Description of SSP System

UML (Unified Modeling Language) is the graphic language of visualization, specification, design and documentation of IT systems. In practice, UML assumes the form of graphic representation of the system being created. This representation consists of the number of logically connected diagrams which enable description of the system, starting from general models up to very detailed ones. UML Standard classifies schemes of diagrams into the following groups: structure diagrams, dynamics diagrams, interaction diagrams and implementation diagrams [11]. Variety of diagrams allow to present the software from many angles and to indicate and emphasize features that are important for a designer in a particular situation [6]. Automatic transition from model being presented on diagrams to low level programming code, e.g. C or Object Pascal is one of the UML language's advantages.

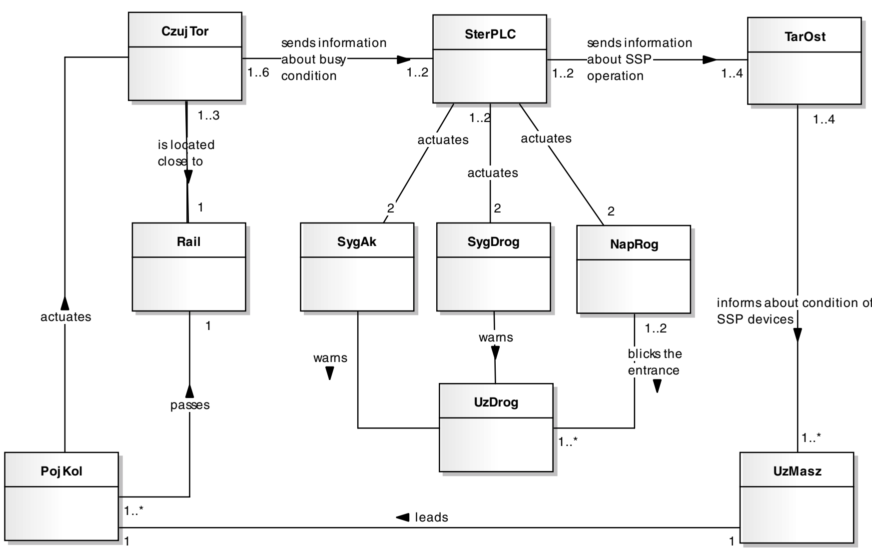


Fig. 2. Class diagram of SSP system

In the Fig. 2, a class diagram is presented in order to exemplify the SSP structure. This is a graphical representation of static elements of the objective domain as well as of connections (associations) between them. The class is a generalization of sets of objects with the same attributes, operations, associations and importance. The example of diagram for SSP system is shown in Fig. 2. On this scheme the most important elements and objects of the SSP system including description of their mutual associations are presented. These are both controlled objects (e.g. road signal devices) and objects supervised within the system (e.g. railway sensors).

The state machine diagram is another type of diagram that can be used for description of the SSP system. What is crucial about this type of diagram is the assumption that during their life cycle objects pass through numerous different stages. This is represented schematically in Fig. 3. In this figure conditions of SSP system from the moment of the beginning (input) to the end (output) of the warning procedure have been depicted. Some of these conditions may be additionally particularized through

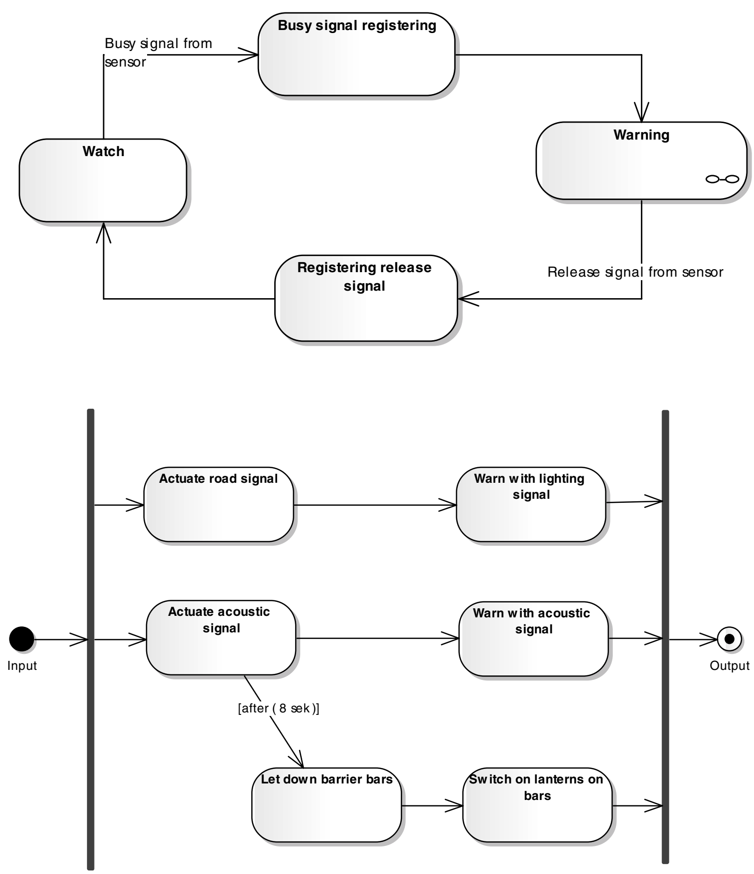


Fig. 3. State-machine diagram of SSP and submachine state of Warning

creating submachine states. In the Fig. 3, the "Warning" status is such a condition. It was specified more precisely in the lower part of the drawing. This is due to the necessity to develop the description of the procedure course which is one of the most essential in the system. Its transfer to submachine state is determined by the need to preserve the main scheme's clarity.

4 Conclusions

According to UIC and CENELEC, at each stage of the software creation process for the fail safe system, from specifying requirements to approving the software, suitable analyses and methods of the software creation should be applied [2]. Except for this obligatory aspect, the system designer is not able to create an error-free software. This is essential because railway traffic control systems are systems conditioned by security and their incorrect operation may jeopardize human life or health. The SSP system is a system which requires highest safety level (SIL4), that is why authors used the example of the automatic computerized level crossing signalling.

In the article it has been stated how specification of correct algorithms for the SSP system should be provided and how to make the implementation in the high level object oriented language (using UML example). The modelling UML language allows modelling of the system by means of a dozen different schemes with both dynamic and static structure. This gives the designer a possibility to look at the problem from many perspectives, which makes the future project more reliable and safe.

References

1. Alagic, S., Arbib, M.A.: Projektowanie programów poprawnych i dobrze zbudowanych. WNT, Warszawa (1982)
2. CENELEC EN 50128, Ralway applications – Communications, signalling, and processing systems – Software for railway control and protection systems (2002)
3. Glenford, J.M.: Projektowanie niezawodnego oprogramowania. Wydawnictwo Naukowo Techniczne, Warszawa (1980r)
4. Jones, C.B.: Projektowanie oprogramowania metodą systematyczną. WNT, Warszawa (1984)
5. Lewiński, A.: The problems of software of safety related computer systems in railway transport applications, Technical University of Radom im. K. Pułaskiego, monograph nr 49, Radom (2001)
6. Szejko, S.: Metody wytwarzania oprogramowania. Mikom, Warszawa (2002)
7. Sumiła, M.: Metoda tworzenia oprogramowania sterującego w systemach sterowania ruchem kolejowym, rozprawa doktorska. Politechnika Warszawska, Warszawa (2007)
8. Tarnai, G., Schneider, E.: Formal metods for railway operation and control systems. In: Symposium FORMS 2003, Budapest (2003)
9. Trzaska, K.: Analysis of software reliability and correctness in railways computer systems. In: Zeszyty Naukowe Politechniki Śląskiej, Z.48, Gliwice (2003)
10. Wrycza, S., Marcinkowski, B., Wyrzykowski, K.: Język UML w modelowaniu systemów informatycznych. Wydawnictwo Helion, Gliwice (2005)
11. Zabłocki, W.: Modelling of station systems of railway traffic control. Zeszyty Naukowe Politechniki Warszawskiej seria Transport z.65, Warszawa (2008)

A Measurement of the Information Value in Transport Processes

Kornel B. Wydro[1,2]

[1] University College of Business and Technology in Warsaw
[2] National Institute of Telecommunications,
Szachowa 1, 04-849 Warsaw, Poland
k.wydro@wste.pl

Abstract. The problem of the cost/benefit evaluation at ITS implementation stands for one of the key questions during the system planning. As all the intelligent transport systems are based on the intensive exploitation of information, the question of information value is of a significant meaning. One approaches the problem in a various manners, mostly yet by calculation of the economic effects "post factum". But this method is of low convenience and usefulness, as it may rely mostly on the experience, which brings always some errors resulted by the solutions and circumstances uniqueness. In the paper there are discussed evaluation approaches based on subjective value assessment and probabilistic decision model, in which a value of information is measured as the effect of it influence on decision choice, with considering the probabilities of a given decision circumstances. Obviously, this method have to be applied "ante-factum".

Keywords: Information, information value, intelligent transport systems.

1 Introduction

A vehicular means, proper infrastructure, energy and information are the indispensable factors for all the transport processes realization. Usually, undertaken transport tasks need decision where and how to transport something in the best possible manner. It obviously depends on the possessed means, and among them, just information play the critical role. In the control systems theory terms all such tasks can be described as optimization task, very often multi-criterial one. The quality of the task solution is measured by the assumed criteria fulfilling level, when in the case of transport, the solution itself consists of the route and transport means selection with consideration of the existing and possible circumstances, mostly occurred on the route during transport time. Therefore, the accessibility of related information and its quality in great part determines the achieved effect, being measured with help of the given utility indexes.

Usually with reference to the quality of the executed transport task, one or a combination of few criteria have to be applied, from which the most important are minimization of the cost, time or risk.

The circumstances taken into account at transport process requirements determination concern transport means and route choice, as well as the conditions on the accessible routes. While vehicles and routes are selected from some disposable determined

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 210–217, 2010.

sets of possibilities, the conditions on routes, mainly weather, air pollution and condition of road surface are only partially known and this knowledge has probabilistic character. So conditions parameters can be taken into account only in probabilistic form, especially, when some calculations have to be done. As all those information are more complete and correct, better decision can be taken regarding the realization manner of the undertaken or already continued transport task. Having in mind a huge number of the transport processes carried out in country scale and its economic meaning, the problems of the proper information collection and usage become very important factors [1], [2]. Furthermore, it justifies a question how to measure value of the information, taking into account – among others – the fact, that the process of acquiring information itself create some costs associated rather with quantity and quality of information, but not with content, and it is reasonable at low information utility to give up the search for it.

Looking generally at the problem, it has to be emphasized that information as an economic resource has some specific features [3]. Firstly, it is not a self-contained resource: it is useful only in case when it helps to achieve some goal in a certain way, and real utility cannot be stated before using it. In that case, the measure of information value in the broadly understood meaning of the usefulness is the grade in which it facilitates or improves the process of the given aim achieving. Secondly, information is not wearable (cannot be used up): even multiple use does not even degrade it, but it can usually become worthless with time. And thirdly, the usage, concerning the manner and level, depends on the assumed goals and technical and mental capabilities of the user. Consequently, the value of information has to be determined subjectively and in reference to the value of the process[1] in which it can be applied [13], [11]. If the criterion of the process become of monetary cost, then the reduction of this cost can be the basis for a measure of the used information value. Obviously, at other process evaluations criteria, not monetary, even information value will be properly measured in other entities, but there always exists possibility to search for some recalculations to monetary entities or construct some standardized, universal indexes[2].

2 Basic Methods of Information Value Measurement

A survey can be stated from the issue of the attempting cases that in the area of information evaluation two basic methods may be applied:

- Direct subjective value assessment,

and

- Evaluation derived from decision process in which the information is used.

Below both the methods are shortly described.

2.1 Direct Subjective Information Evaluation

The often applied approach, presented by economists define the (economic) value of information in the context – as it was mentioned above – of an optimal choice problem, based on common observation that an potential user is making a choice to maximize

---

[1] The value of the process effect is subjective.

[2] Such a normalized indexes are necessary at multi-criterial depictions.

the expected utility or minimize the expected cost of acquired and applied information. In that approach the value of information is the increment in the expected utility resulting from the improved choice made thanks to better information, and often this can be translated into some monetary equivalent representing how much someone would pay to acquire a given piece of information. It is a method known as WTP/WTA one, which describes a difference between Willingness To Pay and Willingness to Accept Compensation (shortly – Willingness to Accept) WTA) [4], [5]. It seems that WTP/WTA method fits better to transport process value than to information value determination, though in some cases can be applied just to evaluate the information. In the formal sense, acquiring of proper information for an effective molding of decision comes from aspirations to lessen the uncertainty regarding the achieving of the wanted results or possible results increasing. In that sense an information, as it was already said, is not an ordinary consumer's good, but rather a "raw" material used for production of some other goods. As just information has indirect value, manifested at its use, the trials of direct normative information utility value measurement are rather not promising [4], [14].

Even auxiliary decisions, concerning questions of what kind of information may reduce uncertainty, where to search for it, and how much it can be worth, also are burdened with some uncertainties, as hardly ever it is known in advance of what quality the found information will be , in what degree it will reduce the uncertainty bounded with the effect of the decision regarding main goal of activity. All that results from the fact that information is an "empirical good", which value can be known only after its use, and from usual lack of "meta-information" which could be helpful in a-priorical assessment of information itself.

2.2 Decision Theory Model of the Information Evaluation

In decision making analysis, the quality of the decision result is usually measured with so called *utility function*. It is a function describing relation between influencing factors and decision outcome measured in decision-maker's usefulness scale. For illustration of the dependability, on the Fig. 1 is shown an hypothetical utility function and a function describing growth of the costs of acquiring the related information (marginal costs) and a point in which the continuation of the information acquiring becomes unprofitable[3]. Usually in practice, the outcomes of utility function are defined as an ordered set of preferences.

As it was mentioned earlier, it is rational to assess the value of information as the not self-contained good indirectly, in relation to the effects of the taken decision with regards towards some goal, in which the information was an inflicting factor. The decision is therefore based on goal's hypothesis C, e.g. a hypothesis related to the possibility of goal accomplishment considering the information, describing a circumstances having direct influence on the of reached goal quality level. From the mathematics perspective, the correctness of circumstances estimation is dimensioned by the probability measures [8]. Also C is a probabilistic variable with possible values c, and possible decisions (actions) d are represented by decisive variable D , which are deterministic and are chosen by decision-maker from some finished set of possibilities.

[3] Obviously, in real situations such a functions can have quite different forms, and more, identification of this curves is quite difficult.

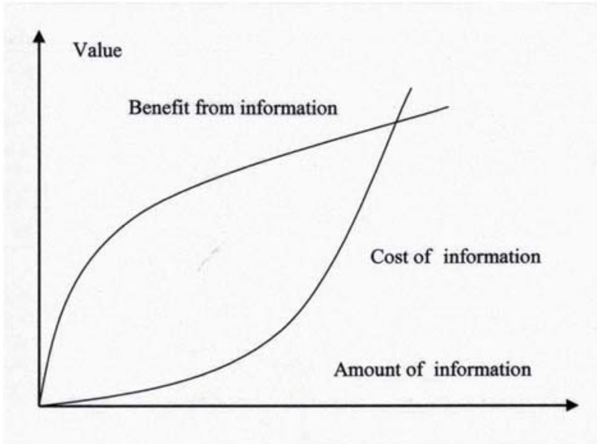


Fig. 1. Hypothetical functions of information utility and acquiring cost

The effects of realization of taken decisions in the given circumstances can be characterized by some values or "utilities" representing a level of desirability (satisfaction) of decision-maker. Generally it may be said that the influence of the circumstance factors *c* and chosen *d* on gained result can be reflected by the utility function *U(d,c)*. an optimal decision is in accordance torational decision making rules , which maximizes expected value of *U(d,c)* at given credibility to circumstances state, i.e. known distribution function of state probability. Also, if there is known the distribution function *p(c)* of random variable *C*, then the expected value of the possible utility value is given by equation (1):

$$\mathbf{E}U(C) = \sum p(c)U(d,c)\,. \tag{1}$$

So consequently, there should be chosen such a decision, which at given state c maximizes gained utility – more precisely – maximizes expected value of it, what can be formally expressed by formula

$$\mathbf{E}U_{\max}(\mathrm{C}) = \max \sum p(c)U(d,c). \tag{2}$$

In many cases the situation of the state C related to goal's hypothesis cannot be observed directly. In such a case it is necessary to conclude on some other observable premises, which can bring any knowledge related to goal hypothesis. In such a case a so called *indicative variable* is used, e.g. some quantity bounded in some manner with goal's hypothesis. Let it be marked by *I*, and it's values by *i*, and its probability distribution as *p(i)*. Assuming that total probability of relation between I and *C* is known, the distribution probabilities of *I* and *C,* as well as conditional probabilities *p(I|C)* and *p(C|I)* can be calculated. In practice the common probabilities may be effectively calculated by the Bayesian networks of likelihood [6][4].

[4] There exists the algorithms for calculation of probability distributions on the basis of common probabilities.

Let us assume that there occurs a need to use some indicative variables, and that a cost of such knowledge gathering involves expenses K_I. A question arises what kind of indicators may be gathered and what part of the reserves may be allocated for it? Just for this aim the utility of information on behalf of goal's hypothesis should be evaluated. A value of information used as indicative variable may be defined as a function of the difference between utilities gained at strategy of action undertaken with consideration of the indicative variable and without it. In the first case the expected value of utility is

$$\mathbf{E}U_m(C|i) = \max \sum p(c|i)U(d,c)\,. \tag{3}$$

As the result of indicative variable consideration cannot be known in advance, it is proper to calculate an expected value of utility as a mean value of all the possible values of I. In this case an expected value of maximal utility will be given by equation

$$\mathbf{E}U_m(C|I) = \max \sum p(i)\,\mathbf{E}U_m(C|i)\,. \tag{4}$$

In the second case it is obviously $\mathbf{E}U_m(C)$, and the information value will therefore be:

$$WI(C|I) = \mathbf{E}U_m(C|I) - \mathbf{E}U_m(C) \tag{5}$$

and considering the cost of information acquiring, the net cost (6):

$$WI_{net}(C|I) = \mathbf{E}U_m(C|I) - \mathbf{E}U_m(C) - K_I. \tag{6}$$

3 A Transport Supporting Information Case

The variety of information usage in transport [17], when it concerns goals of transport processes as well as entities managing those processes, allows both methods discussed above, as well as their applications, for information evaluation. Coming nearer to validation assessment, it is necessary to formulate some general remarks concerning some universal features of information with regards to the elements which have to be considered, when analysing the evaluation problem.

It is worth in this place to mention that from the general socio-economic point of view, the value of information may be considered as a direct value of it, utilitarian value for defined processes, for various social activities in diverse dimensions as economic, cultural, military, etc.

Looking from the discussed research subject point of view, the most important factor for information validation is a set of operational features assessed from the transport task goals, although other dimensions may be taken into consideration. More accurate analysis of the question shows that a proposal of information classification in following two dimension is most legitimated. Those are:

- areas of applications (areas of utility),
- use conditions.

The assumption that the area of application (designation) of information stands for basic classification criterion may be rooted in the primacy of the meaning and role of information in transport (similarly in any case as in each other domain). Such areas in intelligent transport domain are mostly determined by services provided by the given system for which the given information is needed. Obviously, a group of systems having similar service tasks create the distinguished application areas. It is worth mentioning that to some degree, the influence on a given system classification may have also its technical solution applied in specific cases, which are often unique from the construction point of view, but can be quite universally applied in various systems (i.e. vision systems, which may be used for safety improvement, traffic control, vehicle identification etc.).

Due to that, as the systems usually are not one-task ones, even their attribution to application areas are not unambiguous. Similarly unambiguous are the descriptions (names) of applications areas. Currently, the names of application areas are quite stable, which have a good reflection in work division among Technical Working Groups in committees of standardisation institutions, such as ISO and CEN, in research works documentation and publications in ITS area, as well as in professional intercommunication.

In turn, from the formal usage conditions point of view, the gathered and distributed information have to be categorised as:

- obligatory,
- contracted,
- free (unconstrained).

It is strongly related to legal ownership of information, the rights of availability of it, but also with formal regulations concerning the technical means used to gather, distribute and present information [15]. Undoubtedly, it is a very important factor and needs to be taken into account in elaboration and adoption of the rules and standards of information operations.

Moreover, information standardisation rules elaborations shows that information communiqués (announcements) have to be formed according to some ordered structural forms [16] and competed with auxiliary data characterizing a given communiqué (being in fact a meta-information, i.e. information about information). Those necessary processing influences the value of information.

In very common case the entity initiating some transport type processes decide according to its own recognition or experience what kind of information they need or recognize as useful and what expenses they are willing to spend for information acquiring. It is worth to emphasize that the thing may concern the information already existing, for example about road condition on chosen routes or accessible connections in multi-modal journey. Such information is usually relatively cheap and easy to collect. In other case there may be the need to acquire more advanced information, for example the specialized weather forecasts, which have to be properly prepared. In that case the cost of information may be essentially greater. Both those cases are evidently related to WTP/WTA formula. Respectively decision-probabilistic formulas of information validation may be related either to the choice among the transport task realization variants, for example the choice of some preferred route from given set o possibilities, having as a criterion the probability level of appearance of some critical factor, for example the weather on possible routes, or a manner of transport task realization on arbitrarily chosen route. In this second case it is necessary to select or to formulate some measure of task realization quality such as time of realization, cost of required energy or probably most interesting – the risk of the task failure. For calculating the value of information having impact on task realization and being the basis of reaching the extreme of the value of the expected effect, it is necessary to build a countable utility function binding particular values of the utility with various decisions supported by acquired information concerning the probabilities of circumstances of task realization.

4 Conclusions

As problem of information value, the evaluation become a very important factor in contemporary knowledge based economy A transport sector, being strongly dependent on information should be much more carefully analysed. It is the more so, because usually the systems planner and constructors seems to disregard the meaning and value of information. The analysis of the cost/benefit approaches in intelligent transport architectures seems to confirm this thesis. Therefore, the WTP/WTA cases described above and other possible models should be studied more deeply. Furthermore, problems of information standardization, important for effective information management, its ownership providers and users responsibility for information reliability – should be on the research schedule, as the information have to be treated as other market goods.

References

1. Wydro, K.B.: Intelligent Transport Systems - an issues outline. Magazyn Autostrady 3, 72–79 (2009) (in Polish)
2. Wydro, K.B.: Information ows in intelligent transport systems. In: III International Conference Transport Systems Telematics, Katowice (2003)
3. Wydro, K.B.: Information features, basic tools. Wydawnictwo Naukowe Obserwacje, Warsaw (2009) (in Polish)
4. Ahituv, N.: Assesing the value of Information: Problems and approaches. In: Proceedings of ICIS 1989 (1989)
5. Raban, D.R., Rafaeli, S.: Subjective value of information: the endowment effect, University of Haifa (2002)
6. About Bayesian Belief Networks. Charles River Analytics, Inc., Hingham (2004)
7. Hammitt, J.K., Shlyakther, A.I.: The Expected Value of Information and the Probability of Surprise. Risk Analysis 19(1) (1999)
8. Wierzbicki, A.P., Wydro, K.B.: Information Issues of Negotiation. Obserwacje Wydawnictwo Naukowe, Warsaw (2006) (in Polish)
9. Rafaeli, S., Raban, D.R.: Experimental Investigation of the Subjective Value of Information Trading. Journ. of Association for Information Systems 4 (2003)
10. Van Wegen, B., de Hoog, R.: Measuring the economic value of information. Journal of Information Technology (1996)
11. Weissinger, T.: Information as Value Concept: Reconciling Theory and Practice. Working Paper University of Illinois (2005)
12. Sing, A.: ILM: What's your information worth? Network Magazine India (2004)

13. Tappenden, et al.: Methods for expected value of information analysis in complex health economic models: developments on the health economics. Health Technology Assessment 2004 8(27) (2004)
14. Wydro, K.B., Olender-Skorek, M.: Value of information. National Institute of Telecommunications, Warsaw (2006) (in Polish)
15. Wydro, K.B., Gut-Mostowy, H.: eCall system as intelligent vehicle onboard equipment. Magazyn Autostrady 10, 12 (2009) (in Polish)
16. CEN Technical Committee 278, http://www.cen.eu/cenorm/sectors/
17. Mikulski, J.: General Characteristics of the Transport Telematic Systems. In: Telematyka Transportu Drogowego, pp. 73–103. Wydawnictwo ITS, Warsawa (2009) (in Polish)

Application of Shape Analysis Techniques for the Classification of Vehicles

Krzysztof Okarma and Przemysław Mazurek

Higher School of Technology and Economics in Szczecin
Faculty of Motor Transport,
Klonowica 14, 71-244 Szczecin, Poland
{okarma,mazurek}@wste.szczecin.pl

Abstract. One of the main tasks of the statistical traffic analysis is its rating due to the size, type or number of axles. A typical method for measuring the volume of traffic along with the initial classification is based on data derived from inductive sensors and load cells. The possibilities of such a system are however limited, therefore in recent years a great interest in machine vision systems can be observed. An interesting image analysis technique that allows a rapid classification of the types of vehicles observed from the side view is the shape analysis. It can be applied for binary images, for which the values of shape descriptors such as e.g. Feret's diameter can be calculated, as well as some additional quantities such as the center of gravity determined for greyscale images. The article presents the results of the shape analysis obtained for different types of vehicles observed from the camera placed beside the road.

Keywords: Intelligent Transportation Systems, image analysis, shape analysis.

1 Introduction

Classification of the vehicles is usually based on using the built-in road sensors. Such approach has it own limitations related to the way of performed measurements. Since the measured signals are typically obtained from one or two sensors, the amount of data is relatively low. Such classical systems may be sufficient if only the longitude or weight of some vehicles are required but e.g. the height of such vehicle would be often impossible to measure. More advanced classification and analysis requires the use of the machine vision systems. Since the traffic density is often measured just by observers, the automatic classification systems based on relatively cheap equipment would be beneficial, allowing conducting a more comprehensive analysis of traffic.

There are many image analysis algorithms used for the recognition of human faces, register plate numbers etc. Nevertheless, such methods are usually very complex and require relatively high computational power. In this paper the simplified approach for the vehicles' classification based on the analysis of binary images acquired from the side view located camera is analyzed. In such case

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 218–225, 2010.

both the longitude and the height of the vehicles can be estimated, assuming the calibrated camera is located near the road. Nevertheless, even for the uncalibrated camera there are some possibilities of performing a reliable classification of vehicles using the image analysis approach, since the shape coefficients can be used, which are often independent on the object's size.

2 Application of Shape Analysis in Intelligent Transportation Systems

In order to acquire the images, which could be useful for further analysis, the proper location of the camera should be chosen. Depending on the weather conditions, the incorrect observation angle may cause some serious problems related e.g. to the presence of shadows caused by a strong sunlight. For this reason the camera should be mounted relatively low, preferably on the sunny side of the road in order to prevent the presence of the shadows between the vehicles and the camera. Another problem may be caused by the light reflections. The application of simple luminance based thresholding may lead to many artifacts on the resulting image, which may be useless for further processing.

Considering the impact variable lighting conditions as well as the various colors of vehicles, in some cases similar to the background (e.g. dirty dark gray cars), a simple thresholding should not be used. Nevertheless, a more advanced approach requires some additional steps like background estimation and removal so the process of the classification should be as fast as possible. Taking into account the fact that some of the vehicles, especially passenger cars, are similar to each other, shape analysis techniques do not allow a proper recognition of the vehicle's model. For this purpose some other recognition methods would be much more useful, such as register plate number recognition for the comparison with the database. In some of such systems the additional information such as detected average color and indication of the vehicle's brand, usually located centrally over the register plate, may be used. Such systems require the cameras located in front of the appearing vehicles.

In order to perform the shape analysis the background estimation and removal algorithm should be used at first. The simplest approach is the acquisition of the image containing the empty road and its surroundings and the comparison of each acquired video frame with that reference image in order to obtain the binary mask representing the objects which appear on the image or change their locations. One of the disadvantages of such simple method is its sensitivity on changing lighting conditions and some additional motion e.g. trees. Limiting the expected size of the object representing the vehicles such noise introduced e.g. by the wind can be easily eliminated. However, changing lighting conditions require some updates of the background reference which can be performed e.g. on the analysis of the average luminance outside the binary mask representing the vehicle or vehicles. All the update operations should be related to the reference background image (not just the previous frames, since some vehicles may be stopped e.g. by the traffic lights). In the high density traffic conditions the initial background may be estimated using the averaging of the images acquired during a specified period of time [4].

3 Typical Shape Descriptors Used in Image Analysis

For the recognition of different shapes various geometrical features can be determined. Obtained estimation can be used mainly for the reduction of the amount of data for the classification. Such compact description of the image content is used manly for the binary images, which can be obtained from the video sequences using the background estimation and removal algorithms.

The geometrical features, which are the easiest to determine, are the object's area and perimeter, since they can be estimated by counting all pixels fulfilling specified conditions. In such case counting of all pixels representing the vehicle in the binary image leads to the estimation of object's area and the same operation preceded by the edge detection allows the estimation of object's perimeter. For the edge detection various filters can be used, both for binary and grayscale image as well as the color ones. The most typical applications is the morphological contour detection and convolution filters using some typical masks known as Sobel, Prewitt, Roberts and Kirsch filters. Slightly more complicated technique is the application of the Canny filter [1], which has been used in our experiments.

The geometrical parameters used in the image analysis applications can be divided into two major groups: local (such as e.g. the average area or the mean diameter) and global (e.g. the number of objects within the given area). Assuming the presence of a single object on the analyzed fragment of the image plane the estimation of the relative global parameters is unnecessary. Such simplified situation is assumed in our further investigations. for the vehicle's classification purposes the most interesting parameters are those which are insensitive to image deformation's which are possible during their acquisition, as well as some common geometrical transformations (scaling, translation and rotation). In such sense the usefulness of the simplest parameters, such as area or perimeter, is limited, especially for the uncalibrated camera with unknown exact relation between pixel's size and the real dimensions of the objects visible on acquired images.

Some other useful object's geometrical parameters are Feret's diameters being the measure of horizontal and vertical object's maximum size and horizontal or vertical projection's lengths (easily extended by analysis the presence of concavities in the object's shape) [3]. Some more advanced features are linear moments defined typically for binary images e.g. the first order ones corresponding to the object's Center of Gravity as:

$$M_{1x} = \frac{1}{A(X)} \cdot \sum_{X} x_i \ , \quad M_{1y} = \frac{1}{A(X)} \cdot \sum_{X} y_i \ , \tag{1}$$

where X stands for the analyzed object, $A(X)$ denotes its area and x_i, y_i are the coordinates of the i-th pixel. Such moments can also be determined for the grayscale images as the Center of Gravity useful e.g. for the analysis of objects with various colors of the surface. Using the additional edge detection before the calculation of these moments the coordinates of the centroid can be obtained instead of the Center of Gravity.

The second order moments correspond to the object's inertia measures and are defined as:

$$
\begin{aligned}
M_{2x} &= \frac{1}{A(X)} \cdot \sum_X (x_i - M_{1x})^2 \ , \\
M_{2y} &= \frac{1}{A(X)} \cdot \sum_X (y_i - M_{1y})^2 \ , \\
M_{2xy} &= \frac{1}{A(X)} \cdot \sum_X (x_i - M_{1x}) \cdot (y_i - M_{1y}) \ .
\end{aligned} \tag{2}
$$

The most interesting parameters for recognition purposes are those, which can be easily and fast estimated, since the reduction of the computational complexity can be achieved. Such parameters are represented by e.g. the shape coefficient defined as:

$$R_A = \frac{L^2}{4 \cdot \pi \cdot A} \ , \tag{3}$$

where L denotes the object's perimeter and A is the object's area. Another parameter related to the measure of longitudinal character of the object is Feret's coefficient defined as:

$$R_F = \frac{L_h}{L_v} \ , \tag{4}$$

where L_h and L_v are horizontal and vertical Feret's diameters respectively. Circular shape of the object can also be determined by the following circular coefficients:

$$R_{C1} = 2 \cdot \sqrt{\frac{A}{\pi}} \ , \qquad R_{C2} = \frac{L}{\pi} \ , \tag{5}$$

related to the diameters of the circles with the perimeter and area identical to given object's ones respectively.

Another measure is the object's compactness defined as:

$$C = \frac{A}{L_h \cdot L_v} \ . \tag{6}$$

The analysis of the object's shape can be further extended using some more complex techniques, such as e.g. PDH (Point Distance Histogram) shape descriptor [2] calculated using the polar coordinates with the origin placed in the centroid, higher order moments and additional shape coefficients such as e.g. Blair-Bliss coefficient:

$$R_{BB} = \frac{A}{\sqrt{2 \cdot \pi \cdot \sum_i r_i^2}} \ , \tag{7}$$

where r_i denotes the distance of the i-th pixel to the Center of Gravity. A similar descriptor is known as Daniellson's coefficient calculated as:

$$R_D = \frac{A^3}{\left(\sum_i l_i\right)^2}\,, \tag{8}$$

where l_i is the minimum distance of the i-th pixel to the object's contour. Using the image obtained as the result of edge detection another descriptor can be computed, known as Haralick's coefficient:

$$R_H = \sqrt{\frac{\left(\sum_i d_i\right)^2}{L \cdot \sum_i d_i^2 - 1}} \tag{9}$$

where d_i is the distance of the i-th contour pixel to the object's Center of Gravity and L is the number of contour pixels (in some application another number may be used depending on the method of the perimeter's estimation).

Nevertheless, taking into account the possible applications in the low performance and embedded systems more complex shape descriptors presented above have not been considered in our experiments.

4 Experiments and Results

In order to verify the usefulness of some simple shape descriptors, possible for the fast estimation, for the real-time traffic analysis based on the machine vision, a number of images have been tested and the shape coefficients obtained for them have been compared with the values obtained for the synthetic shapes described in [7]. The procedure of the image analysis begins with the background estimation and removal in order to obtain the binary mask for further processing. After that, the noise filtration can be performed using the morphological opening and closing operations as well as the detection and removal of the small objects, which may be present on the images. In the next step, the images have been trimmed in order to increase the calculation of the shape coefficients, so that the longitude of the vehicle is equal to the image width expressed in pixels and the vehicle's height to the height of the image respectively. Such values can be directly used for the calculation of the Feret's coefficient according to (4). In order to obtain a convenient binary contour image, the Canny filter has been used preceded by the extension of the image by the additional background rows and columns (one from each side) in order to prevent missing the boundary contour pixels being the result of the previous trimming. For such prepared binary images some of the shape coefficients have been calculated and the results are presented in Table 1. A modified set of the reference binary images is presented in Fig 1 extended by some additional shapes representing a motorbike and two additional types of passenger cars (wagon and terrain/SUV). The two additional

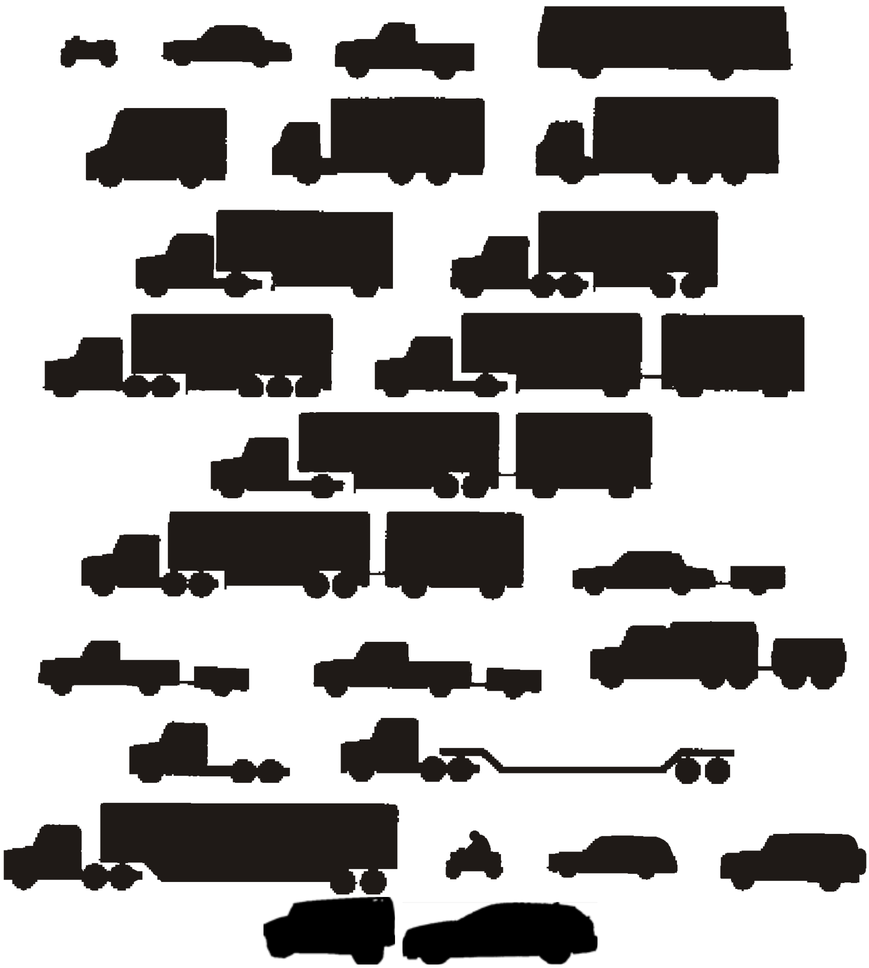


Fig. 1. A set of reference binary images and two exemplary test ones used on the experiments

test images should be classified as these two types of vehicles - the first one as the terrain/SUV and the second one as the wagon.

Analyzing the results presented in Table 1 some interesting relations between the shape coefficients can be noticed. In the last four columns the relative coordinates of the Center of Gravity and the centroid (calculated for the trimmed image) are presented. Sorting the table by each column it can be noticed that the two test images can be classified using the vertical coordinates of the Center of Gravity or centroid, but the use of only these two parameters is not enough for the proper classification. Much better results can be obtained using the compactness or the

Table 1. The values of the shape coefficients obtained by the reference images and two exemplary test images

Type	R_F	$R_S \cdot 10^3$	R_{C1}	R_{C2}	C	$COGx$	$COGy$	Cx	Cy
1	1.8136	0.2470	0.9369	0.0147	0.6894	0.4979	0.5204	0.4802	0.5143
2	3.0625	0.0974	0.9067	0.0089	0.6457	0.5143	0.5304	0.5002	0.5486
3	2.4393	0.0738	0.8770	0.0075	0.6041	0.4933	0.5522	0.5053	0.5491
4	3.4638	0.0293	1.0382	0.0056	0.8466	0.4980	0.4432	0.4962	0.4470
5	1.7483	0.0350	1.0208	0.0060	0.8184	0.5474	0.4882	0.5317	0.5044
6	2.3976	0.0331	0.9987	0.0057	0.7834	0.5501	0.4916	0.4962	0.5047
7	2.7365	0.0297	1.0129	0.0055	0.8058	0.5427	0.4821	0.4934	0.4973
8	2.9571	0.0416	0.9901	0.0064	0.7699	0.5544	0.4997	0.4943	0.5466
9	3.0798	0.0452	0.9854	0.0066	0.7626	0.5613	0.5014	0.5175	0.5616
10	3.3374	0.0479	0.9936	0.0069	0.7754	0.5604	0.4970	0.5232	0.5675
11	4.7633	0.0401	0.9759	0.0062	0.7480	0.5395	0.4715	0.4905	0.5349
12	4.8715	0.0703	0.9306	0.0078	0.6801	0.5391	0.4926	0.4758	0.4590
13	4.9940	0.0388	0.9905	0.0062	0.7705	0.5376	0.4702	0.4940	0.5175
14	4.8072	0.1041	0.8713	0.0089	0.5963	0.4700	0.5545	0.5080	0.5737
15	3.7619	0.0871	0.8043	0.0075	0.5081	0.4548	0.5747	0.5028	0.5830
16	3.9907	0.0779	0.8201	0.0072	0.5282	0.4503	0.5562	0.5031	0.5719
17	3.7874	0.0476	0.9538	0.0066	0.7145	0.4926	0.5234	0.5196	0.5375
18	2.6174	0.0799	0.7951	0.0071	0.4965	0.4043	0.6124	0.4673	0.6459
19	6.0000	0.1276	0.6316	0.0071	0.3133	0.3805	0.6285	0.5072	0.6676
20	4.3333	0.0311	0.9823	0.0055	0.7579	0.5500	0.4672	0.5014	0.5184
21	2.2619	0.0449	1.0002	0.0067	0.7857	0.5447	0.4770	0.5264	0.4899
22	1.7440	0.0509	1.0395	0.0074	0.8487	0.5312	0.4900	0.5245	0.5058
23	1.1758	0.1516	0.8629	0.0106	0.5848	0.5153	0.5925	0.4900	0.5869
24	2.9878	0.0872	*0.9436*	0.0088	*0.6993*	0.5443	0.5078	0.5064	0.5285
25	2.5370	0.0556	**0.9647**	0.0072	**0.7309**	0.5354	0.4845	0.5063	0.4992
26	1.9253	0.0258	**0.9629**	0.0049	**0.7282**	0.5444	0.4746	0.5409	0.4661
27	3.1009	0.0058	*0.9500*	0.0023	*0.7089*	0.5274	0.5139	0.5066	0.5363

circular coefficient R_{C1}. The nearest values of these two shape coefficients for the recognized types and the test images are indicated in the table.

The use of some other shape coefficients, even for the preliminary classification may lead to wrong results, so our future work will be directed towards the development of a fast and reliable classification method based on a combination of some shape descriptors. In order to determine the impact of each coefficient on the shape recognition accuracy, some more comprehensive tests are planned with the use of much larger dataset. For testing purposes the Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) may be helpful for the proper reduction of the dimensionality in the domain of the shape coefficients.

5 Conclusions

The results presented above indicate the possible use of some relatively simple shape coefficients for the fast classification of vehicles useful for the real-time statistical traffic analysis. Such approach can be especially relevant for the existing systems which use low cost cameras, which usually can acquire only a few frames per second with reasonable resolution. Nevertheless, the advantages of the machine vision systems, even simplified, are evident in comparison to the classical measurement techniques based on road sensors and inductive loops.

Some of the shape coefficients analyzed in the paper can also be efficiently estimated using the Monte Carlo approach, especially for the high resolution images [6]. Such approach is based on the analysis of a number of randomly drawn pixels from the image in order to determine its parameters [5] and can also be used for the fast video analysis purposes. The ability of the fast image analysis is especially important in low performance systems, often based on the industrial cameras.

Acknowledgments. This work is supported by the Polish Ministry of Science and Higher Education (Grant No. N509 399136 „Estimation of the vehicles' motion trajectories using the Bayesian analysis and digital image processing algorithms").

References

1. Canny, J.: A Computational Approach To Edge Detection. IEEE Trans. Pattern Analysis and Machine Intelligence 8(6), 679–698 (1986)
2. Frejlichowski, D.: Trademark Recognition Using the PDH Shape Descriptor. Annales UMCS – Informatica 8(1), 67–73 (2008)
3. Kindratenko, V.: Development and Application of Image Analysis Techniques for Identification and Classification of Microscopic Particles. PhD thesis, Antwerp University (1997)
4. Okarma, K., Mazurek, P.: Background Estimation Algorithm for Optical Car Tracking Applications. Machinebuilding and Electrical Engineering 7-8, 7–10 (2006)
5. Okarma, K., Lech, P.: Monte Carlo Based Algorithm for Fast Preliminary Video Analysis. In: Bubak, M., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) ICCS 2008, Part I. LNCS, vol. 5101, pp. 790–799. Springer, Heidelberg (2008)
6. Okarma, K., Lech, P.: Application of the Monte Carlo Preliminary Image Analysis and Classification Method for the Automatic Reservation of Parking Space. Machine Graphics and Vision 18(4), 439–452 (2009)
7. Traffic Detector Handbook: Third Edition, Publication No. FHWA-HRT-06-108, Federal Highway Administration, vol. I, US Department of Transportation (2006)

Additional Warning System for Cross Level

Andrzej Lewiński and Lucyna Bester

Kazimierz Pułaski Technical University of Radom,
Faculty of Transport and Electrical Engineering,
26-600 Radom, Malczewskiego 29, Poland
{a.lewinski,l.bester}@pr.radom.pl

Abstract. The paper contains an analysis of the safety level crossing equipped with an additional warning system for drivers that are within the level crossing before the approaching train. The proposed system is based on wireless data standard, WiMax and sensor networks WSN, placed an additional warning helps to improve safety at unguarded railway crossings. Mathematical analysis was carried out for unguarded level crossing model, and then for system with signaling the level crossing ssp and for system equipped with additional warning system for drivers. For the analysis presented models used stochastic Markov processes which allowed estimating the indicators of probabilistic studied systems.

Keywords: Safety of railway cross level, Markov process, wireless transmission, warning and protection system.

1 Introduction

Almost all accidents at unguarded crossings are the result of mistake or imprudence of drivers. Therefore, in order to eliminate or at least limit this type of accident model was developed wireless warning system for drivers at level crossings. In this solution a data radio module based on WiMax technology and sensor networks WSN was introduced. The presented system can warn road users located in the area of railway crossing, before the approaching train [3] [5]. The paper contains the mathematical analysis of three models of the level crossings: unguarded railway crossing, the level crossings equipped with the automatic protection system APS (barriers and signaling), and proposed a level crossing equipped with an additional warning system for unguarded crossing. It allows estimating of the required probabilistic measure and carrying out the comparative analysis of models.

2 Estimating the Level of Safety for the Railway Crossing

The main measure used to characterize the safety railway system is the probability of dangerous situation. For the analysis the homogeneous, ergodic and stationary Markov processes are assumed for modeling the real behavior of road users in railway crossing area. Movement of vehicles at the level crossing is described by Poisson process, characterized by the exponential distribution of time of events. This method gives a possibility to estimate the probabilities of dangerous situation [6].

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 226–231, 2010.

2.1 Modeling Events at Railway Crossings

Fig. 1 shows the simple model of unguarded level crossing equipped with warning signs only. Bellow the following states describing a typical situation are defined:

0 – main state, the cars pass over the railroad tracks, the lack of presence of a train

1 - represents the state to stop before the railway crossing,

2 - state in which the driver did not stop before the crossing and tried to drive through the crossing,

P4 - catastrophic situation in which driver drove into the railroad tracks before the approaching train.

In the graph the following transitions are introduced: λ1 – intensity of transition from state 0 to state 1, λ2 - intensity of transition from state 0 to state 2, λ4 -intensity of transition from state 0 to state 4 (the driver drove into the oncoming train), λ5 - intensity of transition from state 2 to state 4 (the driver did not stop and drove into oncoming train), μ1 – inverse of the waiting time before the level crossing, μ2 – inverse of the time crossing the railway track, μ4 – inverse of the time return to the basic state.

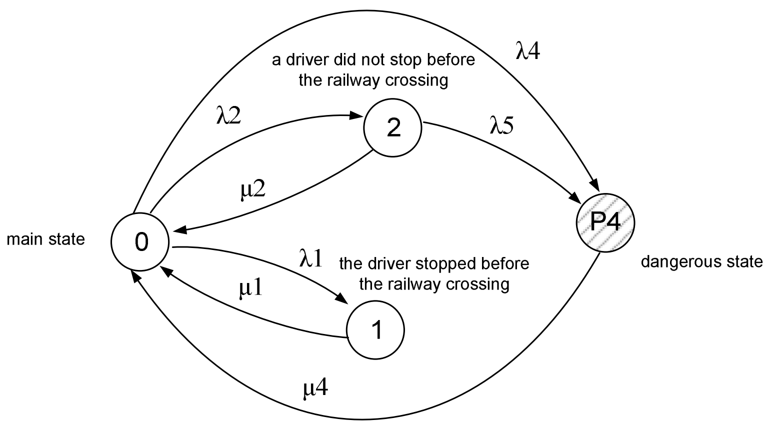


Fig. 1. Model of the behavior of traffic participants at unguarded railway crossing

Corresponding to this model the concept of a model railway system with additional early warning drivers before approaching train is proposed. In this model, an additional system sends radio messages, to the driver informing about train in intersection area [3] [6]. In relation to the previous model states 1 'and 2' were introduced as follows:

1'- state in which the driver received a message about the train and stops before the railway crossing,

2'- state in which the driver received a message warning of the train and did not stop before the railway crossing.

The transitions are introduced: λ1' – intensity of transition from state 0 to state 1 (the driver received the message and stopped before the railway crossing), λ2' - intensity of transition from state 0 to state 2 (The driver received the message and did not stop before the railway crossing).

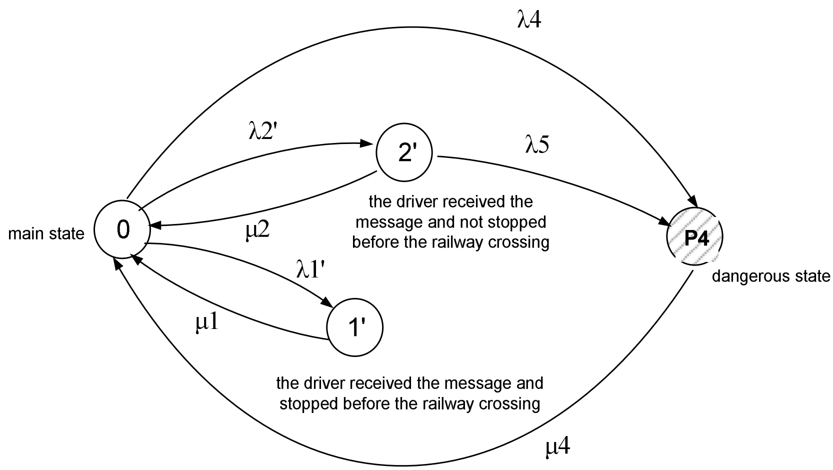


Fig. 2. Model of the behavior of participants in traffic on unguarded railway crossing with an additional warning system for drivers

Fig. 3 presents the model of railway crossing equipped with a system of automatic signaling the level crossing. The following description of the states was introduced:

1”- the state describing the turn on of the railway warning system, the driver stops before the railway crossing,

P4- catastrophic situation in which driver take part in collision with train caused by signaling system fault railroad tracks before the approaching train.

The following transitions are defined: λ1” – intensity of transition from state 0 to state 1 (the driver stopped before the railway crossing), λ4” - intensity of transition from state 0 to state 4 (the driver did not stop before the level crossing and drove into oncoming train due to failure of the APS)

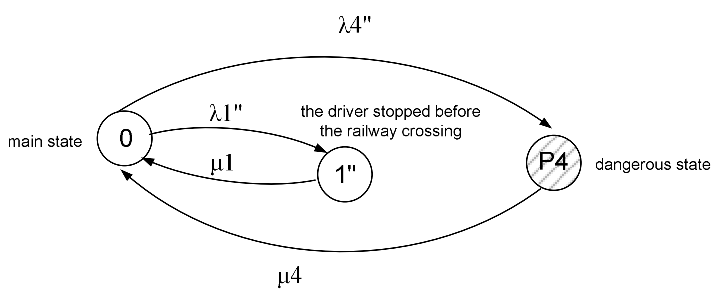


Fig. 3. Model the behavior of traffic participants on a level crossing equipped with automatic signaling system of level crossing

2.2 Analysis of the Occurrence of Dangerous Situations

The safety analysis of presented models requires the estimation of the probability of dangerous state P4. In all models, a similar intensity of arriving vehicles at the railway crossings was assumed. The parameters assumed in Table 1, 2, 3 reflect typical traffic situations (f.e. 5 min. between cars corresponds to $\lambda = 12\ h^{-1}$) and general assumptions that for arrows outgoing from state 0, the sum $\Sigma\ \lambda_i$ for all models it must be the same.

For the model unguarded railway crossing as shown in Fig. 1, the probability of catastrophic state of P4 is:

$$P4 = P4(t)_{t\to\infty} \frac{\mu_1(\lambda_2\lambda_5 + \lambda_4(\lambda_5 + \mu_2))}{\lambda_2\mu_1(\lambda_5 + \mu_4) + (\lambda_5 + \mu_2)(\lambda_4\mu_1 + (\lambda_1 + \mu_1)\mu_4)} \tag{1}$$

The Table 1 shows the estimated value of P4 for a typical intensity (volume) of the vehicles. The values of parameters corresponding to the description (intensities λ and μ) are related to typical experimental results [6].

Table 1. Assumed values for the model on Fig. 1

5 min. between vehicles	10 min. between vehicles
λ1 = 10	λ1 = 4
λ2 = 1.00012	λ2 = 1.006
λ4 = 0.0000012	λ4 = 0.000006
λ5 = 1.00012	λ5 = 1.0006
μ1 = 20	μ1 = 20
μ2 = 1200	μ2 = 1200
μ4 = 1	μ4 = 1
P4 = 0.0005554089903851569	P4 = 0.0007024634023730639

Similarly, for the model system equipped with an additional warning system for drivers, the probability of P4 may by estimated.

$$P4 = P4(t)_{t\to\infty} \frac{\mu_1(\lambda_2'\lambda_5 + \lambda_4(\lambda_5 + \mu_2))}{\lambda_2'\mu_1(\lambda_5 + \mu_4) + (\lambda_5 + \mu_2)(\lambda_4\mu_1 + (\lambda_1' + \mu_1)\mu_4)} \tag{2}$$

The results are presented in Table 2.

Table 2. Assumed values for the model on Fig. 2

5 min. between vehicles	10 min. between vehicles
λ1' = 12	λ1' = 6.0053
λ2' = 0.00012	λ2' = 0.0006
λ4 = 0.0000012	λ4 = 0.000006
λ5 = 0.00012	λ5 = 0.0006
μ1 = 20	μ1 = 20
μ2 = 1200	μ2 = 1200
μ4 = 1	μ4 = 1
P4 = 7.50006890613031 $^{-7}$	P4 = 4.6146516311914545 $^{-6}$

However, for the model equipped with automatic signaling system APS, the probability of dangerous state P4 is:

$$P4 = P4(t)_{t \to \infty} \frac{\mu_1 \lambda_4 "}{\lambda_4 " \mu_1 + (\lambda_1 "+\mu_1)\mu_4} \quad (3)$$

and estimated value is shown in Table 3.

Table 3. Assumed values for the model on Fig. 3

5 min. between vehicles	10 min. between vehicles
λ1" = 12.00024 λ4 " = 0.0000012 μ1 = 20 μ4 = 1	λ1" = 6.006 λ4 "= 0.000006 μ1 = 20 μ4 = 1
$P4 = 7.499938125510464^{-7}$	$P4 = 4.61419202421528^{-6}$

The intensity of dangerous failure for the signaling system at railway crossings with warning λ4 "is assumed with respect to the [4].

3 Conclusion

The paper proposes a method for improving safety at unguarded railway crossings equipped with an additional warning system based on new information technologies, WiMax and WSN (messages about the train are sending to the drivers in cross level area, these message may by displayed on the special signal board or in the receivers inside cars).

The analysis of the safety level for different models shown in Fig. 1, Fig. 2 and Fig. 3 (models: the unguarded railway crossing, railway crossing with an additional warning to drivers and railway crossing with APS system) allows to estimate the probability of a dangerous situation P4. The probability of a dangerous situation for unguarded railway cross is rather high ($P4 = 5.5^{-4}$) in comparison with application of additional warning ($P4 = 7.5^{-7}$). Such solution may by compared to the cross level equipped with APS system ($P4 = 7.4^{-7}$) when a low intensity of cars is assumed.

The obtained values of probabilities of dangerous states P4 are related to the intensity of traffic, ride the trains, the number of events and an efficient transmission system of warning messages. These results show that the introduction of an additional warning system at unguarded railway crossings has important influence on improving the safety of road users and the additional messages about approaching train can enforce appropriate driver behavior. The results indicate that the proposed system must be classified as a safe system (THR ratio as for SIL 4).

References

1. Beibei, W., Indranil, S., Matolak, D.W.: Performance evaluation of 802.16e in vehicle-to-vehicle channels. In: 66th IEEE Vehicle Technology Conference, VTC Fall 2007. IEEE, Boltimore (2007)
2. Jadźwiński, J., Ważyńska-Fiok, K.: Bezpieczeństwo systemów. PWN, Warszawa (1993)
3. Lewiński, A., Bester, L.: Application of wireless transmission systems for improving safety at railway crossings. Konferencja Transport XXI wieku (2010) (referat przyjęty do druku)
4. Perzyński, T.: Problemy bezpieczeństwa sieci komputerowych stosowanych w sterowaniu ruchem kolejowym – rozprawa doktorska. Wydział Transportu i Elektrotechniki Politechniki Radomskiej, Radom (2009)
5. Lewiński, A., Rogowski, A., Bester, L.: The modeling of collision with train - car respect to integrated transport safety system. In: Mikulski, J. (ed.) Advanced in Transport Systems Telematics, pp. 179–184. WKŁ, Warszawa (2009)
6. Zastosowanie nowych standardów sieci bezprzewodowych w systemach w sterowaniu i zarządzaniu w transporcie lądowym. Instytutu Automatyki i Elektroniki Politechniki Radomskiej. Praca naukowo – badawcza 2604/47/P. Radom (2009)

Problems of Safety Codes Evaluation in Practical Applications

Mária Franeková and Karol Rástočný

Department of Control and Information Systems, Faculty of Electrical Engineering, University of Zilina, Univerzitná 1, 010 26 Žilina, Slovakia
{maria.franekova,karol.rastocny}@fel.uniza.sk

Abstract. The paper deals with a methodology of the safety codes evaluation used in the safety-related communication systems. The problems of determining the probability of undetected error of block detection codes generally used in praxis are mentioned in comparison with the theoretical knowledge. The main part is oriented at the description of the mathematical apparatus for determination of the residual error probabilities of the safety block codes. The practical part describes the results of modelling the failure effect cause by Electromagnetic Interferences (EMI) to safety of the closed industrial transmission system.

Keywords: Safety related communication system, safety code, safety integrity level, probability of undetected error, modelling.

1 Introduction

In the case, where a communication system is a component part of an electronic system which participates in control of a safety critical process, the system has to be designed to guarantee the required safety integrity level (SIL) according to [1]. Safety related (SR) communication must be based on safety principles and principles of COTS (Commercial Off The Shelf) communication technology cannot be used. At present the number of safety related communication technology vendors increases. The questions of SR communications system analysis and synthesis are described in detail e.g. for the area of railway control systems or industrial control systems. In these areas the requirements for safety codes used in the process of data transmission are described in standards EN 501 59-1 [2] (for a closed transmission system), EN 501 59-2 [3] (for an open transmission system) and in the standard IEC 61784-3 [4] valid for the industrial network used for the measurement and control systems.

To keep off the integrity of a message which is corrupted in consequence of Electromagnetic Interferences (EMI) in the transmission channel it is recommended to use, without a transmission code (implemented in the standard layer of communication protocol), a safety code which is situated in the additional safety layer. Safety codes are one of very important techniques in SR applications. The first aim which we regard within the selection of safety codes is maintaining the data integrity but the safety code can be used for data authentication too. In the process of quantitative analysis of SR communication protocol it is necessary to determine the safety features

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 232–242, 2010.

of safety code, mainly the failure probability of the code. A general model of SR message transmission across a closed transmission system with the safety and transmission code is illustrated in Fig. 1.

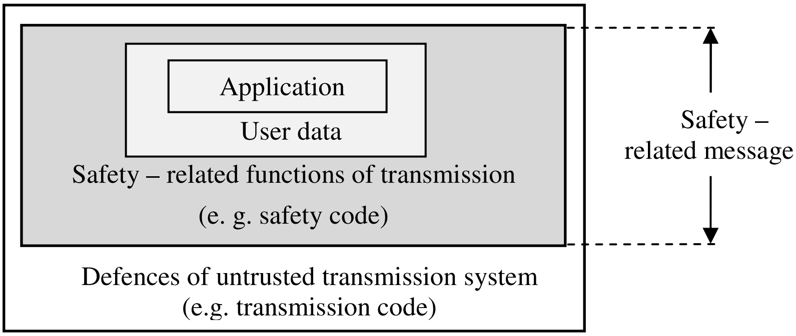


Fig. 1. Model of safety related message in the closed transmission system

2 Problems of Safety Analyses of Safety Codes in Praxis

As it is well known two parameters are used in the safety evaluation process of block detection codes: minimal Hamming distance of code d_{min} and probability of undetected error of code words with lengths n. Majority of the presented results are connected to certain concrete class of codes (e.g. Hamming codes, RS codes,…) and the probability results of undetected error of decoder p_e are valid for certain situations and simplified conditions, which are not applicable in the practice. The next problem is that the presented theoretical proceedings are valid for a certain construction length of a code, which in most cases does not comply with the telegrams lengths used in practice [5]. This is why some relations for p_e determination are very often unusable because of difficult calculation methods, too (especially for large code word length). For this case parallel processing methods and special computation tools may be used [6]. This is why the safety analysis of the safety codes used in practice is very often difficult and we must come out from pessimistic estimation of the probability of corrupted messages only, so called upper estimation. The probability of undetected error of code word p_e depends on bit error probability p_b of the transmission channel used. In practice the different types of transmission channels have different values of p_b and different behaviour of the noise in concrete media. In many cases the transmission channel testing is not possible. The transmission channel can affect the transmitted messages by noise, interference or by fading of signal. These effects are generally referred to as the EMI and they have a significant influence on the value of corrupted messages intensity. The EMI results from different influences, which cannot be described by deterministic relations. The relations for the determination of undetected error probability of a code assume the occurrence of EMI (replacing of one symbol of transmitted message by another symbol) only. Within monitoring of EMI effects for the total value of the intensity of failure there is a tendency to maximally approximate the analyzed situation by a mathematical expression. During the determination of p_e the statistical values of bit error probability in typical transmission media are very often used.

When we know the value of p_b (finding by testing in real conditions or as a statistical value) it is very important to determine p_e in dependence on p_b for all telegrams lengths used in the communication protocol and to prove the monotonicity of function for p_b from 0 to 0.5. In many cases the function is not monotone, what can affect the hazard and the occurrence of dangerous states [7]. For some types of telegram lengths the value of p_e very often exceeds the upper estimation.

The next problem within the real situations is the determination of the number of redundant bits in the safety code for keeping the required safety integrity level. Very often an idea prevails that the bigger number the better safety features, while the other parameters are not maintained and it is not true. In some applications another supplementary procedure very often increases the safety of transmission (e.g. a double assurance of message in direct and inverse directions of message symbols by two different codes).

2.1 Mathematical Apparatus of Bit Error Probability Determination

Bit error probability p_b introduces the error probability of transmission bit „0“ or „1“, which occurs on the receiver side in a detector. The symbol is received as „1“, if its received value is bigger than the defined threshold level. In the other cases the detector evaluates incorrect symbol as „0“, i.e. the bit error occurs. The bit error occurrence is caused by the noise in a transmission channel, which occurs when the value of the received signal is above or under the threshold level, contrary to the expectations. The error occurs if:

$$y_1 + N_S \langle a \quad \text{or} \quad y_0 + N_S \rangle a, \tag{1}$$

where y_1 is the actual occurrence of symbol „1“, y_0 is the actual occurrence of symbol „0“, N_S is the actual value of signal in a defined time and a is the defined threshold level. Then the bit error of the transmitted symbol „1“ is defined by conditional probability (2) and the bit error or the transmitted symbol „0“ by conditional probability (3):

$$p_{b1} = P(0|1) = \int_{-\infty}^{a-y_1} p_{N_S}(N_S|y_1)dn, \tag{2}$$

$$p_{b0} = P(1|0) = \int_{a-y_0}^{\infty} p_{N_S}(N_S|y_0)dn. \tag{3}$$

After the application of Gaussian channel model to (2) and (3) for the probability of occurrence of Gaussian noise applicable to symbols y_0 and y_1 we will get:

$$p_{N_S} = p_{N_S}(N_S|y_0) = p_{N_S}(N_S|y_1) = \frac{1}{\sigma_{N_S}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{N_S}{\sigma_{N_S}}\right)^2}. \tag{4}$$

After arranging expression (4) it is possible to determine the probability of symbols „1“ and „0“ as:

$$p_{b1} = P(0|1) = \frac{1}{2} erfc\left(\frac{y_1 - a}{\sigma_{N_S}\sqrt{2}}\right), \tag{5}$$

$$p_{b0} = P(1|0) = \frac{1}{2} erfc\left(\frac{a - y_0}{\sigma_{N_S}\sqrt{2}}\right). \tag{6}$$

Function *erf*(x) is an error function defined by

$$erf(x) = \frac{2}{\sqrt{\pi}} . \int_0^x e^{-t^2} dt. \tag{7}$$

Function *erfc*(x) is the complementary error function. The threshold level a is such that the consequential probability of error increases. Then the consequential bit error probability is

$$p_b = P(0|0).p_{b0} + P(1|1).p_{b1}. \tag{8}$$

The minimal value of bit error occurs, if:

$$P(0|0) = P(1|1) = \frac{1}{2}. \tag{9}$$

For the threshold level valid

$$a = \frac{y_0 + y_1}{2}. \tag{10}$$

After connection with previous relations the consequential bit error probability p_b of transmission channel, which characteristic represents an additional Gaussian noise is

$$p_b = \frac{1}{2} erfc\left(\frac{y_1 - y_0}{2\sqrt{2\sigma_{N_S}}}\right). \tag{11}$$

2.2 Mathematical Apparatus of Probability of Undetected Error Determination

Some approaches to determination of residual error probability [8], [9] need to know all code words of the code. Then the probability of undetected sequence error p_{e1} for Binary Symmetric Channel (BSC) can be calculated according to (12):

$$p_{e1} = \sum_{i=\left\|\frac{d_{\min}+1}{2}\right\|}^{n} A_i \, p_b^{\,i} (1 - p_b)^{n-i}, \tag{12}$$

where d_{min} is the minimal Hamming distance of code, A_i is the total number of sequences in the code words with weigh of i, p_b is the bit error probability of channel.

A few classes of safety codes for which the complete weight function of code words $A(x)$ is known only, where $A(x)$ is the weight-enumerating function of a code:

$$A(x) = \sum_{i=0}^{n} A_i x^i. \tag{13}$$

This class includes e.g. the binary linear Hamming perfect (n, k) codes, Reed - Solomon (RS) codes over Galois field GF(q).

The weight-enumerating function for the distance-3 Hamming codes of code words length $n = 2^r$-1 (where r is number of redundancy bits) is [10]:

$$A(x) = \frac{1}{n+1}\left[(1+x)^n + n(1+x)^{\frac{n-1}{2}}.(1-x)^{\frac{n+1}{2}}\right]. \tag{14}$$

The weight-enumerating function of Reed-Solomon codes with q symbols can be calculated:

$$A_0 = 1, A_i = 0 \qquad \text{for } (0 < i < d_{min}),$$
$$A_i = \binom{q-1}{i}(q-1)\sum_{j=0}^{i-d_{\min}}(-1)^j\binom{i-1}{j}\left(q^{i-d_{\min}-j}\right) \qquad \text{for } (d_{min} \le i \le n). \tag{15}$$

Both Hamming and RS codes are very often used in a safety related communication protocol as the safety codes. Hamming codes based on polynomials generating are equivalent with cyclic codes used in the polynomials generator. It is necessary to point out, that relation (14) is valid for original construction length of Hamming code and not for shortened codes. The Reed-Solomon codes are MDS (Maximum Distance Separable) codes in which the safety characteristic of shortened MDS code is the same as in the code within original construction length [6].

The weight structure of non perfect safety codes can be determined in two ways:

- Direct calculation - for the short code words length only.
- Calculation using the dual codes (so called Mac Williams's identity [11]) - for the code words of large length computational is very difficult and very often needs to have efficient computational tools or parallel processing of data.

For this reason we very often use a simplified relation for the determination of the residual error probability or the maximal value of estimation $2^{-r.}$ For the block safety codes (n, k) with code word of lengths n and with generation of polynomial $g(x)$ equation (12) can be modified by (16), in which the value of A_i is approximated by (17).

$$p_{e2} \cong \frac{1}{2^{n-k}} \sum_{i=d_{\min}}^{n} \binom{n}{i} p_b{}^i (1-p_b)^{n-i}, \tag{16}$$

$$A_i \cong \frac{1}{2^{n-k}}\binom{n}{i}. \tag{17}$$

If np_b<<1 in (16) then, the sum can be approximated by the first term of sum (18):

$$p_{e3} \cong \frac{1}{2^{n-k}}\binom{n}{d_{\min}} p_b{}^{d_{\min}} (1-p_b)^{n-d_{\min}}. \tag{18}$$

It is evident that in expression (18) besides parameters of n, k it is necessary to know also the minimum Hamming distance of code words $d_{min.}$ If this value is unknown the Gilbert's inequality for even length of code words (19) and for odd length of code words (20) can be used for determination of d_{min}.

$$2^k \sum_{i=0}^{\frac{d_{\min}-1}{2}} \binom{n}{i} \leq 2^n , \tag{19}$$

$$2^k \sum_{i=0}^{\frac{d_{\min}-2}{2}} \binom{n-1}{i} \leq 2^{n-1} . \tag{20}$$

3 Modelling of Failure Effect on Safety of Closed Transmission System

Let's assume the end to end communication system (Fig. 2). This communication system consists of safety-related equipment SRE 1, SRE 2 and a trusted transmission system with safety profile ProfiSafe [12], which realises safety-related functions within transmission in compliance with [1]. The base of the trusted transmission system includes a non-trusted transmission system (Profibus DP), which insures transmission messages by transmission code TC. To achieve the required safety level of transmission, transmission messages have to be ensured by safety code SC. It is necessary that the encoder and decoder of safety code are realized on a fail-safe principle. A transmission channel, which is influenced by electromagnetic interference, is a component part of a transmission system. Authors assume only a closed transmission system and encoders/decoders independence of safety and transmission codes, too.

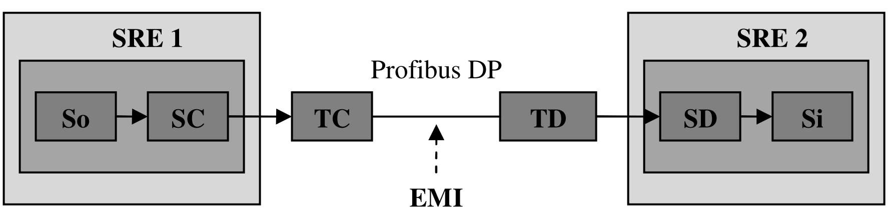


Fig. 2. Closed transmission system with the transmission and the safety code

Safety related messages are generated in safety related source So of SRE1 and after the transmission received by the safety related sink Si of SRE2. Transmission decoder TD checks the correctness of the message transmitted, which can be corrupted during transmission by EMI. If the transmission decoder evaluates the message received as a fault the system goes to the defined state. This state depends on the technology applied and the specifications of communications not-related to safety. In most cases a repeated message transmission is required. If the situation is opposite, when the message received is evaluated as correct, the message is transmitted across the communication chain to safety decoder SD. If the safety decoder does not detect a communication fault the message is transmitted to the sink of information. If the message is evaluated by the transmission decoder as a fault the system goes to the defined state. According to the standard, the message correction is not allowed as a reaction to fault. Generally, the request of message is transmitted or the message is transmitted in a defined period (in a cycle).

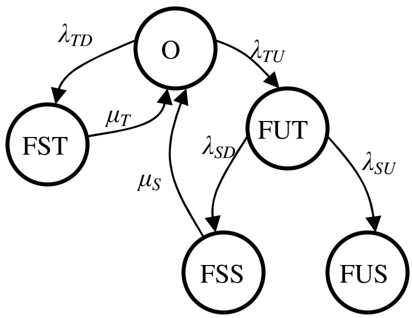


Fig. 3. Markov chain of the industrial transmission system with transmission and safety codes

The coincident effect of several factors on the transmission system safety can be demonstrated using a Markov chain. Fig. 3 illustrates the Markov chain of the transmission system from Fig. 2. The model consists of five states. The meaning of particular symbols in the diagram in Fig. 3 and the description of the transition in diagram are illustrated in Table 1 and Table 2.

Table 1. Descriptions of states in the diagram illustrated in Fig. 3

State	Description of states
O	The transmission system is functional. Transmission of uncorrupted messages between pieces of equipment.
FST	The system state when the transmission decoder detects a corrupted message. The state is concerned as a fail-safe state.
FUT	The system state when the transmission decoder does not detect a corrupted message. The state is concerned as a fail-unsafe state.
FSS	The system state when the safety decoder detects a corrupted message. The state is concerned as a fail-safe state.
FUS	The system state when the safety decoder does not detect a corrupted message. The state is concerned as a dangerous state.

At the beginning the transmission system is in an operating state (O), in which the safety related equipment change off the messages. After detection of the communication fault the system goes to a fail state caused by the transmission code (FST). If the transmission code does not detect the message, the system goes to a fail-unsafe state caused by the transmission code (FUT). This state is a dangerous state for the system. If an additional safety code detects a fault, the system goes to a fail-safe state caused by the safety code (FSS). The state is defined by a defined reaction to the fault, from which it goes to the operating state (O). If the additional safety code does not detect a fault message, the system goes to a fail-unsafe state caused by the safety code (FUS).

Table 2. Descriptions of transitions in the diagram illustrated in Fig. 3

Transition	Description of transitions	Meaning of transitions intensity
O→FST	The transition is realized in consequence of a transmitted message failure, which is detected by the transmission code.	λ_{TD}
FST→O	The transition is realized in consequence of the mechanism which reacts to a failure (e.g. by a cyclic transmission) and the transition goes again to the operating state.	μ_T
O→FUT	The transition is realized in consequence of a in the transmitted message, which will be not detected by the transmission code.	λ_{TU}
FUT→FSS	The transition is realized in consequence of a transmitted message failure, which was detected by the safety code.	λ_{SD}
FSS→O	The transition is realized in consequence of a mechanism which reacts to the failure (e. g. with operator's acknowledgement) and the transition goes again to the operating state. In most cases the state depends on acknowledgement from the operator.	μ_S
FUT→FUS	The transition is realized in consequence of a transmission message failure which was not detected by the safety code.	λ_{SU}

4 Results of Safety Analyses

The results of the safety analyses of the transmission system illustrated in Fig. 2 were obtained using a Markov chain which was realized via SW SHARPE [13]. The results are valid for the following input data and conditions:

- The length of telegram of safety not related message is 244 Byte.
- The length of a safety related message telegram is 128 bits.
- The applied transmission code is CRC 16.
- The applied safety codes are CRC 16, CRC 24, CRC 32 (in Table 3 simulation 1, 2, 3).
- The transmission and the safety codes are independent (they use different generated polynomials).
- The value of Profibus DP (medium RS 485) bit error probability was used according to statistical results of 10^{-5}.
- In the model we assume a cyclic mode of operation between the safety-related pieces of equipment.

According to the assumptions mentioned the frequency of transmitted messages is f_m = 18,000 to 21,600,000 messages per hour. The result of critical failure rate λ_{CRIT} undetected by the safety and transmission code in the transmission system can be determined according to:

$$\lambda_{CRIT} = f_F \cdot p_{e\_SC} \cdot p_{e\_TC} , \quad (21)$$

where f_F is the frequency of corrupted messages in the receiver part of the transmission system per hour (we assume a pessimistic situation in which all generated messages are fault), p_{e_SC} is the probability of undetected error of the safety code and p_{e_TC} is the probability of undetected error of the transmission code. The probabilities of undetected error of the transmission and the safety codes were determined according to relation (16).

The intensity of renovation μ_S of the system after the failure detection by the safety code was determined according to the relation:

$$\mu_S = \frac{1}{MTTR_S} , \quad (22)$$

where $MTTR_S$ is mean time to failure of the system from the state in which the system goes after a failure detection by the safety code. The reaction to failure after detection by the safety code is related to the acknowledgement from the operator.

We assume $MTTR_S = 1$ hour. After applying the input data to the model illustrated in Fig. 3 the results of the critical failure rates of messages and the value of mean time failure of system for three simulation cases are illustrated in Table 3.

Table 3. The results of a pre-Markov chain simulation via SHARPE

Meaning of symbols	Simulation 1	Simulation 2	Simulation 3
	TC:CRC16 SC: CRC16	**TC:CRC16 SC:CRC24**	**TC: CRC16 SC: CRC32**
f_m [h^{-1}]	21600000	21600000	21600000
f_F [h^{-1}]	21600000	21600000	21600000
p_{e_TC}	$1,52588.10^{-5}$	$1,52588.10^{-5}$	$1,52588.10^{-5}$
p_{e_SC}	$1,52588.10^{-5}$	$5,96046.10^{-8}$	$2,32831.10^{-10}$
μ_T [h^{-1}]	10800000	10800000	10800000
μ_S [h^{-1}]	1	1	1
$MTTF_{CRIT}$ [h]	$7,62.10^{4}$	$1,95 .10^{7}$	$5,11 .10^{9}$
λ_{CRIT} [h^{-1}]	$1,31197.10^{-5}$	$5,12524. .10^{-8}$	$1,95816. 10^{-10}$

5 Conclusions

The safety related communication systems are typically resisting against hazardous faults. The failure effects on the communication system can be determined directly by monitoring the original system installation, by simulating the system operation using its model, or by computing or theoretical reasoning. It is necessary to remark that strictly safety requirements for the safety-related communication system are not achievable by tests only or by results from practice (frequency of dangerous states occurrence is very low and the mean time between failures multiply the excess value of useful lifetime of one safety-related communication system). The aim of analysis of the failure effects on system safety is to form a model, which allows identifying the system transition process from the safety state to a dangerous state and permits to calculate the probability of dangerous state occurrence of the system as a failure effect on the operating system. The transmission system normally does not work in isolation but it is a component part of another superior system, for which it provides a service. Therefore the starting moment of safety model generation is at exact definition of the interface between the transmission system and superior system with the aim to facilitate total identity of threats, which should be considered in the process of analysis. Also it is necessary to explicitly define an event in the safety system output, which is considered dangerous (undesirable) with regard to safety features of the transmission system. Generally, a violated transmission data is considered an undesirable event, which is not detected by the transmission system and further data is regarded as correct. The knowledge of failures and errors attribute of the transmission system forms basic assumptions for the implementation of measures used not only for failures prevention, but also for errors detection and negation of failure effects within their occurrence. When establishing quantitative safety analyses of a transmission system with the use of safety and transmission codes, the biggest problem is the determination of undetected errors probability of codes for concrete length of telegram and degree of assurance used in praxis. For large code words with not validated code length or parameters of (n, k), an exact expression of undetected error probability is difficult and computationally demanding and we must very often use the approximation for the worst estimation 2^{-r}. The process of safety analyses described in the paper for a concrete example of closed safety-related industrial transmission system can be used after the modification of input parameters for the safety analyses of another type of the end to end safety related communication system, too.

Acknowledgement. This paper was supported by the scientific grant agency VEGA, grant No. VEGA-1/0040/08 „Mathematic-graphical modelling of safety attributes of safety-critical control systems“.

References

1. IEC 61508: Functional safety of electrical/electronic/programmable electronic safety-related systems (1998)
2. EN 50 159 - 1: Railway applications: Communication, signalling, and processing systems – Part 1: Safety-related communication in closed transmission systems. CENELEC (2001)
3. EN 50 159 - 2: Railway applications: Communication, signalling, and processing systems – Part 2: Safety-related communication in open transmission systems. CENELEC (2001)
4. IEC 61784-3: Digital data communications for measurement and control. Part 3: Profiles for functional safety communications in industrial networks (2007)
5. Fujiwara, T., Kasami, T., Kitai, A., Lin, L.: On the Undetected Error Probability for Shortened Hamming Codes. IEEE Transaction on Communications 33(6), 570–574 (1985)
6. Karná, L., Klapka, Š., Harlenderová, M.: Quantitative Assessment of Safety Codes. In: Proceeding of international conference FORM/FORMAT 2008, Budapest, Hungary, pp. 249–255 (2008)

7. Dodunekоvá, R., Nikolova, E.: Sufficient Conditions for Monotonicity of Undetected Error Probability for Large Channel Error Probabilities. Problems of Information Transmission 41(3), 187–198 (2005)
8. Cheong , S. K., Hellman, M. E.: Concerning a Bound on Undetected Error Probability. IEEE Transaction on Information Theory, 235–237 (1976)
9. Dodunekоvá, R., Dodunekov, S.K.: Sufficient Conditions for Good and Proper Detection Codes. IEEE Transaction on Information Theory 43, 2023–2026 (1997)
10. Clark, C.C., Cain, J.B.: Error - Correcting Codes for Digital Communications. Plenum Press, New York (1988)
11. Nikolova, E.: A Sufficient Condition for Properness of a Linear error Detecting Code and Its Dual. Mathematics and Mathematics Education. In: Proc. 34th Spring conference of the Union Bulgarian Mathematicians, pp. 136–139 (2005)
12. TÜV Automotive GmbH, TÜV SÜD Group Electronic Systems, TÜV Product Service GmbH: Profibus specifications. PROFIsafe – Profiles for Failsafe Technology, Report No.: PK55299T, Revision 1.2, Order No.: 700 43831(2005)
13. Sahner, R.A., Trivedi, K.S., Puliafito, A.: Performance and Reliability Analysis of Computer Systems. Kluwer Academic Publishers, Dordrecht (1996) ISBN 0-7923-9650-2

Economic and Social Aspects of Applying Biodiesel Fuel in Road Transport

Mihaela Bukljaš Skočibušić, Natalija Jolić, and Zdravko Bukljaš

Faculty of Traffic and Transport Sciences, University of Zagreb, Vukelićeva 4, 10000 Zagreb, Croatia
{mihaelab,natalija,zdravko.bukljas}@fpz.hr

Abstract. The world trend in automotive industry represents the improvement of the existing vehicle power plants and their further development as well as the use of various alternative fuels. Such tendencies should not be considered only from an entirely technical aspect, but also from the economic, social and strategic aspects of the modern society. In this sense it is necessary to give priority to biodiesel fuel. The production of biodiesel fuel has to be developed in compliance with the increasingly severe exhaust emission standards in designing and realization of road transport means. From the economic aspect at macroeconomic level, the development of biodiesel will reflect on the condition of industrial production, employment, additional inflow of financial means into agriculture and the economic development of rural areas, as well as the foreign currency reserves of a country along with the reduction in the dependence of macroeconomic parameters on the external factors.

Keywords: vehicles, biodiesel, traffic policy, economic factors.

1 Introduction

At the end of the last century, i.e. in the 1970s and 80s, there were several crises related to fossil fuels that had economic and political consequences resulting from the fast development of technology with significant increase in the use of fossil fuels. Due to supply and prices of fossil fuels there have been disturbances in the total national economies of many countries. However, greater use of fossil fuels has resulted also in sudden increase in environmental pollution, which along with awareness about the need for sustainable management methods, as well as problems that have occurred in industry and agriculture, has brought to the necessity of developing and using renewable sources of energy.

Biodiesel is renewable, bio-degradable, and high-value liquid motor fuel. It is obtained from rape oil or other vegetable oils (e.g. palm oil) by estherification with methanol, as well as from animal fat and recycled waste edible oil. This process produces a fuel of propulsion characteristics of the same or very similar properties to those of conventional diesel fuel, which is obtained by crude oil refining, but it is standardized liquid non-mineral fuel, non-toxic and bio-degradable. In this way, biodiesel as propulsion fuel with such properties can completely substitute diesel fuel, but can also be mixed with it. With biodiesel-diesel fuel mixture, in which biodiesel

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 243–252, 2010.

accounts for more than 20%, the harmful emissions are significantly reduced. This fact gives powerful stimulation for the usage and production of biodiesel due to the increasingly severe environmental standards that require reduction of greenhouse gas emissions.

2 Biodiesel Application

Solving of problems regarding fuels for internal combustion engines can be considered through two different conceptual approaches. One is finding new engine designs with the possibility of using diverse fuels, and the other represents the application of new and renewable sources of energy. In this sense, the application of biodiesel fuel in the world produces significant economic and strategic possibilities of development.

The use of biodiesel in the world has been going on for many years, but serious implementation and promotion in EU has started to develop more significantly as late as in the last decade. A strong reason for this is the EU policy regarding reduction of carbon dioxide (CO2) emissions in compliance with the Kyoto Protocol and the acceptance of fact that biodiesel does not emit additional amounts of CO2, except during the transport process and the biomass production, and the finished product. The proof for this lies in numerous studies that have shown that the use of 1kg of biodiesel leads to the reduction of the emission by approximately 3kg of CO2 in relation to fossil diesel, considering the entire process.

Biodiesel in internal combustion engines can be used in two ways: as addition to pure diesel fuel, i.e. as addition to fossil fuel in a certain ratio (so called blending); and as pure biodiesel.

Due to relatively low production and absolutely high consumption of fossil fuels, and almost no difference in retail price between pure biodiesel and eurodiesel (e.g. in Germany the difference in retail price amounts to about 5 euro); biodiesel is usually used in combination with fossil fuel, i.e. by blending B5, B20, B50 and B100, where B denotes biodiesel, and the number denotes its share i.e. methyl ester in the mixture. By adding methyl ester the lubrication is improved, not increasing at the same time the sulphur content, whereas in diesel fuels of fossil origin its lubrication properties are reduced by reducing the sulphur content. This means that high lubrication of biodiesel in comparison with mineral diesel causes lower wear of pistons, sealing rings, cylinder walls and precision parts of high-pressure fuel injection pumps. Conventional diesel engines use without any problems the fuel with 20% biodiesel, and many newly designed engines can also use pure (100%) biodiesel. Therefore, biofuels do not require manufacture of completely new automotive engines, since already the existing diesel engines are in this sense compatible.

2.1 Advantages of Applying Biodiesel in Road Motor Vehicles

It is a known fact that great amount of cargo and people transport is performed by road which almost completely depends on fossil fuels. The most commonly used fuels are diesel and eurodiesel. Similarly, if the economically less developed countries in Europe are considered, one can notice that the consumption of motor fuels is constantly increasing, in compliance with their traffic activity.

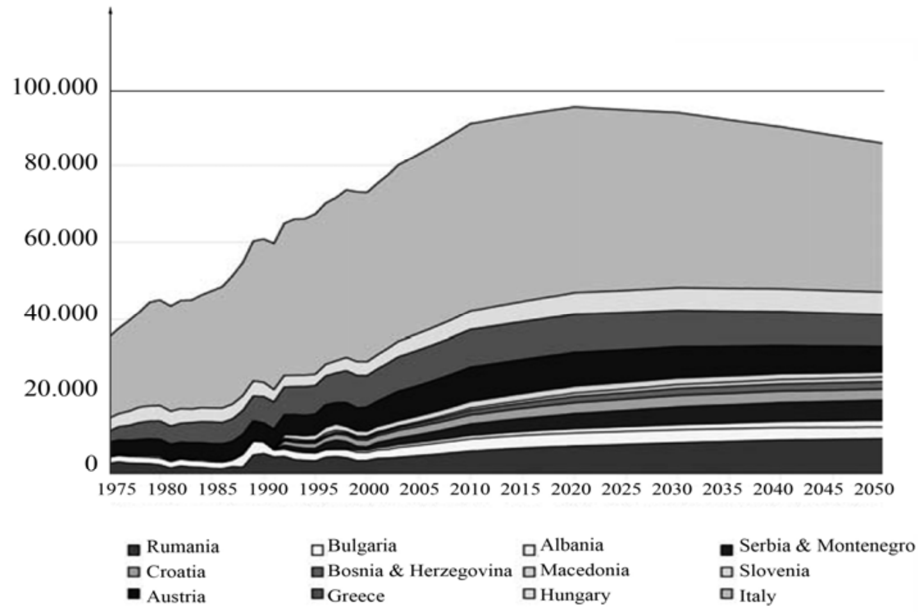


Fig. 1. Motor fuel consumption forecast. Source: [6]

Contrary effect regarding the amount of motor fuel consumption is the constant increase in the efficiency of other transport modes that should result in absolute reduction of the consumption of motor fuels after 2025.

The use of biodiesel, compared to fossil diesel, is suitable considering the environmental protection since it reduces the greenhouse gas emissions and other pollutants. The quantification of the negative effects on the environment is performed through a popular approach "Well-to-Wheels" (WTW) which measures the net emissions during the entire production and consumption chain. However, WTW emission result can significantly deviate from case to case, depending on the very production process and the use of by-products. The basic advantage of using biodiesel as renewable fuel is to significantly reduce CO2 emissions, with the reduction of sulphur oxide emissions, airborne particulates and carbon monoxide. The level of biodiesel efficiency depends significantly on the ratio of the mixture used, as well as engine operation i.e. engine type. The values of potential reduction of individual pollutant emissions during biodiesel use, and based on WTW basis, are the following:

- Every tonne of fossil diesel releases about 2.8 tonnes of CO2 into the atmosphere, whereas the specific carbon content of one tonne of biodiesel is somewhat lower and amounts to about 2.4 tonnes CO2. However, if one assumes that this carbon will be completely used during the next crops that will provide raw materials for the production of vegetable oil, as well as absorbed through the carbon cycle as glycerol and solid waste, then it may be considered that net CO2 emission during use of biodiesel almost equals zero.
- Today, in EU, a tonne of conventional fossil diesel contains on the average maximally 350ppm (parts per million) of sulphur. In this way, during diesel

fuel combustion apart from nitrogen oxides (NOx) also sulphur oxides (SOx) are emitted into the atmosphere contributing to the formation of the so-called acid rains. Biodiesel contains almost no sulphur, i.e. its content amounts to only 0-0.0024ppm.

- Nitrogen oxide emission from biodiesel can be increased or reduced in relation to the emission from fossil diesel, depending on the power engine generation and the very procedure according to which they are tested. The emission of nitrogen oxides from pure biodiesel is increased on the average by about 6% compared to fossil diesel. Regarding the lack of sulphur in biodiesel it is possible to use the nitrogen oxide control techniques that cannot be used in case of fossil diesel.
- Biodiesel contains oxidants that improve the combustion process and reduce emission, which means also the emission of carbon monoxide.
- Breathing in of airborne particulates has proven to be a serious problem and danger for the health of people. Emission of such particulates in exhaust gases in case of using biodiesel is by about 40% lower than in case of fossil diesel.
- In case of possible leakage into the environment, fossil diesel is then only 50% degradable, whereas in the same situation biodiesel is 98% degradable without any further harmful consequences.

Tests have shown that diesel engines that use pure biodiesel B100 have significantly lower emission of smoke and particulate matter (PM). This reduction amounts to about 40%. Lower smoke and particulates emission is achieved e.g. in case of mixtures of conventional diesel and methyl ester (e.g. for B20 the reduction in particulate emission amounts to 12%, carbon monoxide 12%, and hydrocarbons 20%, but the nitrogen oxide emission is increased by 2%). An important data is that B100 reduces the risk of carcinogenic illnesses caused by hydrocarbons by 94%, and B20 by 27%.

The nitrogen oxide emission in case of diesel engines using biodiesel is higher since the methyl ester molecules contain chemically bonded oxygen. However, the emission of polycyclic aromatic hydrocarbons (PAH) is lower by about 80%.

Since the exhaust emissions generated by transport means is significant and an ever increasing source of pollution, the governments in the world are trying to introduce such politics that will contribute to reducing the amount of harmful combustion products from automotive engines, and one of the possible solutions is certainly the substitution of fossil fuels by biological fuels.

2.2 Tendencies in EU Policy for Use of Biodiesel in Road Traffic

Increased harmful gas emissions that participate in the greenhouse effect due to increased demand and use of fossil fuels, and their impact on the global climate, are the reason for the attempts to reduce the emissions of such gases, especially CO2. In the context of Rio de Janeiro Conference and the Kyoto Protocol, EU has accepted the obligation to reduce by 8% the total amount of harmful greenhouse gas emissions with high CO2 content by the year 2012, compared to the year 1990. Consequently, the EU Council accepted in June 1998 the European Commission strategy on adopting the plan about usage of renewable sources of energy with special emphasis on biofuel and brought a respective decision. Thus, the European Commission has defined the Green Paper which stipulates the strategy of energy supply and determines the methods of substituting the fossil fuels by biofuels in the transport sector by 5.75% until 31 December 2010. This then means that according to the EU 2003/30/EC Directive the share of biofuel consumption should grow at a rate of 0.75% per year.

Should biodiesel be used only in automotive engines then it has to satisfy the technical standards according to EN 14214 standard of the European Committee for Standardisation (CEN). Similarly, the EU Council agreed in Göteborg in 2001 on a common strategy for the development of measures for the introduction of biofuel in the EU based on the White Paper which defines the reduction of CO2 emissions in transport sector by 50% in the period from 1990 to 2010, which represents a reduction of 1113 million tonnes of CO2. A high 84% of this refers to road transport.

In accordance with the Brussels Directives from January 2008, EU has set as its goal that the share of biofuels in the fuels used for transport should be increased by as much as 20% by the year 2020. This, among other things, is to insure the stability of investments in this sector and to increase the resulting revenues considering primarily the environmental protection. The transport sector uses over 30% of total energy consumption and is almost completely dependent on the current fossil fuels. Therefore, the policy of transport is the priority area of improving the energy and ecological efficiency.

3 Economic Aspects of Using Biodiesel

Biodiesel is being widely produced in the majority of European countries and has been recognised as efficient solution of the interwoven fields such as agriculture, energy industry, environmental protection and economic and social policies. EU is the most important producer of biodiesel on the global market. Since 1993 the biodiesel production in EU has increased almost ten times, with Germany being the leading EU country, followed by France, Italy and the Czech Republic. Across EU biodiesel has been used in automobiles with diesel engines in mixtures of different ratios with fossil diesel. In Germany, Austria and Sweden pure biodiesel is used in public transport vehicles modified to this type of fuel.

According to the European Biodiesel Board data, the production of biodiesel in EU occupies currently about 1.4 million hectares of fertile land, and there are some forty processing plants which produce annually about 3,184,000 tonnes of this fuel. The number of countries with biodiesel production is continuously increasing, with proportional increase in the capacities for biodiesel production allowing thus further expansion of this type of production in the EU. However, out of the total marketed biodiesel about 12% is imported from Southeast Asia in the form of raw palm oil.

One such example comes from Bulgaria with the recent data of the economic analysis of the University of Roussou for the Rousse region, which show the advantages of biodiesel production and provide the information that 100 tonnes of produced biodiesel can result in 8 million profit tax (BGL) of incomes and can create a thousand new workplaces. Research in the area of agriculture and more efficient production will obviously reduce the costs and the price of biodiesel in the future. Therefore, it is important to use the future potentials in introducing new technologies in manufacturing internal combustion engines with lowest possible harmful emissions, and to increase the possibility of using energy equivalent engines with zero emissions.

Table 1. SWOT analysis

S	*W*	*O*	*T*
1. Low price / exemption from taxes	1. Limited supply of vehicles using biodiesel for B30-B100	1. Issue of global / local warming	1. Availability of biodiesel products
2. Domestic resources	2. Too few petrol stations	2. Long-term increase of oil prices	2. Stable legislation and fiscal system
3. Low cost i.e. adaptation of petrol stations			3. Wide public demand
4. Possibilities of combining different ratios			4. Competitive forced ignition engines

The simplest presentation of production and use of biodiesel to prove this can be done through a SWOT analysis.

Generally, the main advantages in planning and production of biodiesel are the macro-economic conditions on the market of energy sources and the changes in the supply of fossil fuels depending on their availability and reserves. Uncertainties on the market of energy sources are caused by the changes both in energy and environmental policy, including the liberalization of the market of primary energy sources and electrical energy, as well as the intensity of climatic changes caused by the huge amount of greenhouse gas emissions.

The needs to introduce and use biodiesel depend on the economic development and activities of individual countries, i.e. on its possibilities for gradual transition to biodiesel vehicle propulsion, with the possibility of using more economical, more efficient and environmentally friendly vehicles, by introducing more efficient public transport, and introducing special fuel tax based on the carbon content in the fuel.

The economic aspect of production and usage of biodiesel can be considered from the positive and the negative standpoint.

From the positive standpoint at the macro-economic level the biodiesel production development would be caused by increase in employment rate by opening new workplaces, increase in the industrial production, introduction of the so-called “third cultures” (apart from wheat and corn) which would ensure additional and safer revenue to agricultural producers, by using fertile as well as neglected agricultural areas, by allowing better usage of machinery, increase in the cost-effectiveness of agricultural production and economic development of rural regions, by increasing foreign currency reserves of a country, by supplementing by renewable fuel the part of fossil fuels that are imported, and by higher safety in the supply of energy through diversification of energy sources and suppliers. This assumes that the very development of complete industry would be enabled by progressive attitude of each country regarding the policy of subsidies, tax policy, as well as long-term strategy in energy resources management.

However, from the negative standpoint of the economic aspect it should be noted that the use of food for the production of biodiesel increases the price of raw materials on the world stock markets, and that eventually this can lead to world hunger due to the increase in food prices, but this area penetrates the social aspect. It should be noted that several years ago the temporary balance in the global demand and supply of cereal crops was disturbed due to the expansion of food investment into biodiesel which resulted in increase in food prices. Should this trend continue, humanitarian organizations warn of far-fetching consequences for the global population. International economists and experts point out that the entire situation with the price increase will not see any significant changes in the next several years. An especially worrying fact is that food, since the established food price index in the far 1845 is the most expensive until now, i.e. that the food price in the last three years has increased by 75%. Since the biodiesel production is in fact direct transformation of food into the propulsion fuel, which could cause additional demand for certain types of food, and thus raise also its price, has stimulated many sociologists and humanitarian organizations to warn of social aspects of producing and using biodiesel, whose negative consequences have already been felt by numerous poor countries.

4 Social Aspects of Using Biodiesel

The production of biodiesel is closely connected with plants that are used for food, and therefore there is justified worry about the impact of the global price and the availability of food. However, the problem is not only in the price, but also in the data provided by Jean Ziegler, UN special rapporteur and sociologist, that 280kg of corn can be sufficient amount for annual diet of a single child. This amount can produce 49 litres of biodiesel, which an off-road vehicle can consume on the average for 200km of travelled distance. This data causes real worry from the social aspect and a desire to immediately stop further production of biodiesel, except in the segment in which biodiesel is produced from recycled oil, or out of plants not used for food, i.e. out of agricultural waste.

An example of such production and usage of biodiesel from edible oil collected from households, restaurants, hotels and similar can be found in Graz, Austria. Oil is collected by the municipal company workers ÖKOSERVICE GmbH which has an active employment programme, and the produced biodiesel is used as fuel in public transport buses. The usage of biodiesel in public transport means in major cities significantly reduces the harmful emissions into the environment.

However, the social aspect of applying biodiesel reaches even further. There is, namely, danger also from harmful effects on the living environment. If the pasture grounds and forests are supplemented by big plantations of one of the plants required for biodiesel production, the biological diversity will be disturbed by reducing the plant and animal species in a certain eco-system. More intensive agriculture requires also greater amounts of irrigation which could certainly in some parts of the country endanger also the water resources. Furthermore, greater application of various additives that improve the yields can bring to increased emission of nitrogen oxides from the soil which form greenhouse gases, and equally by clearing rainforests which are necessary to stop further global environmental degradation in order to enlarge the fertile areas. Additional harmful emissions would be created by the combustion of fossil fuels during cultivation, transport and processing of plant material necessary for the production of biodiesel.

As adequate alternative to the production of biodiesel as well as sustainability of producing sufficient quantities of food, the experts have been offering over the recent years various “solutions”. For instance, liberal economists advocate greater market liberalisation and reduction of subsidies to farmers for the production, which eventually results in the increase of tax, decline in quality, and food being more expensive. On the other hand, adversaries of capitalism and free trade emphasise that this will result in an even stronger difference among the living standards of the citizens. The latter alternative refers to genetically modified seeds (GMO) which, according to advocates are classified as a more resistant alternative to natural disasters, soil pollution and actual price increase. However, apart from potential threat to the consumers because of the yet insufficiently known long-term effect on organisms, as a strong argument against GMO, the environmentalists mention also the possibility of their omnipresence in food, as e.g. in Argentina and Brazil, where, due to sudden biodiesel production thousands of hectares of forests were cleared, and they substituted the shortage of food production obtained naturally by GMO production. The third alternative, which seems almost the most acceptable results from the use of green algae which live in waste waters and sea, and require very little sunlight and carbon dioxide. Today, in the world, dozens of laboratories are involved in research of obtaining biodiesel from green algae. Green Fuel Technologies from Cambridge in U.S. Massachusetts is one of the leading companies that have developed the process of using algae located in plastic bags in order to collect carbon dioxide emitted by power plants. Algae have the capability not only of reducing the greenhouse gas emissions, but of destroying other pollutants as well. Some types of algae produce starch that can be converted into ethanol, and others produce drops of oil that can be turned into biodiesel or even into fuel for aircraft propulsion engines, and in favourable conditions they can double the mass in a few hours only, as pointed out by the head of the project Marcus Gay.

What seems to be a devastating fact in the EU is the desire to introduce penalties for the countries with insufficient production of biodiesel, i.e. generally biofuel. Thus, e.g. EU, which already has millions of biodiesel users, has formed specific legal frames for the production of biodiesel. On the one hand, member countries are given by means of subsidies tax exemption for the production, and on the other hand penalties are introduced for those member countries whose companies supply fuel without adding a certain percentage of biodiesel to standard diesel. In this way the EU attempts to maximally stimulate the production and usage of this type of fuel.

5 Conclusion

Generally, the traffic systems in the last decades were increasingly oriented to road transport, both in passenger and in cargo transport. Such tendency affects the structural changes within the traffic system which mean increase in energy consumption and increase in energy intensity of the traffic sector. Numerous technologies and sources of energy can be used in *locus standi* objects, i.e. factories, whereas the possibilities for transport sector are rather limited by the use of internal combustion engines. In this context, liquid biofuels are the only renewable source of energy which can be used without changing the present vehicle technology. Since biodiesel is significant for automotive sector, many laboratories and research centres in Europe and the world are acting in this field. However, the problem is reflected in the fact that although acting in similar tendencies, they occupy different aspects and different approaches regarding production and use of biodiesel. Obviously, the Brussels guidelines are meant to significantly reduce the greenhouse gas emissions. However, by greater use of biodiesel it is obvious that a part of agricultural areas, that had been used to produce food, will be reallocated to produce raw materials to obtain biodiesel, which could lead to a misbalance regarding the food and ecological chain.

From the economic viewpoint, at the global level, every country is specific regarding the share of biodiesel so that it is necessary to distinguish between the industrially developed countries and the developing countries. The industrially developed countries have the economic power of development, research, improvement and promotion of production and use of biodiesel through different world organizations and programmes, whereas the developing countries are attractive due to their cheap labour. Countries in transition can plan their future depending on the political will and speed of entry into the OECD (Organization of Economics Co-operation and Development), as well as on other associations present in the developed countries.

However, from the social aspect, biodiesel defined as perfect substitute for oil and global salvation from greenhouse gases pollution, has proven as a big threat for the population and the entire eco-system. First, it was found out that due to its production, forests and rainforests as natural filters are being destroyed, and after that came the devastating claim by UN that biodiesel has been bringing hunger and increasing the price of food raw materials. Due to higher profit, many agricultural producers have decided to sell their yield as raw material for fuel production, instead as raw materials for food production. The basic objectives of EU to increase the share of renewable sources of energy in the total energy consumption by the end of this year, have been subjected to serious revision due to the suspicions in the advantages of using biofuel, i.e. in the considered case, the biodiesel. Germany is one of the countries with the highest production of biodiesel in the world, and in the recent year's biodiesel as alternative fuel has become number one in France, Austria, the Czech Republic and the Netherlands as well. However, the scientists today point to yet another fact, which is that during the processing of food raw materials a big quantity of by-products remains (e.g. rape-cake or corn residue) which can also be used in the production of biodiesel. This fact should be obviously accepted as challenge by the agronomic and other professions. This would then fully meet the EU requirements, and through the placement of by-products make the production of biodiesel more cost-effective. It is therefore necessary through integral studies to find the optimal method of production, ranging from the quantities, technology, logistics to the selection of raw materials, everything in relation to the clearly defined agricultural areas that are used, or will be used for this purpose, in correlation with rationally determined needs for energy of the respective origin. Only in this way will it be objectively possible to establish a balance between the technical-technological, economic-social and food-ecological systems at all levels.

References

1. White Paper European Transport Policy for 2010: Time to Decide, European Commission (2001)
2. Institute for Prospective Technological Studies: Techno-economic analysis of biodiesel production in the EU (2002)

3. Energy consumption, European Environment Agency (2003)
4. Analysis of the Impact of High Oil Prices on the Global Economy, International Energy Agency (2004)
5. Kurevija, T., Kukulj, N.: Global environmental issues concerning large scale biodiesel production, Faculty of Mining, Geology and Petroleum Engineering. University of Zagreb, Zagreb (2006)
6. Granić, G.: Region and Croatia in the Concept of WEC Scenario. In: 16th forum: Energy future in the light of the relations and integration processes in Europe, Hrvatsko energetsko društvo, Zagreb, pp. 11–34 (2007)
7. Joel Jr., K.B.: Biogoriva: blagodat ili beskoristan pothvat (Biofuels: blessing or useless venture?). National Geographic Hrvatska 48(10), 10–29
8. Šimičević, H.: "Ekološka" alternativa za naftu i fatalna opasnost za gladne ("Ecological". alternative for oil and fatal threat for the hungry), Nacional, Zagreb (2008)
9. Baraka, B.L.: Biodizel – regulativa i smjerovi proizvodnje (Biodiesel – regulations and production directions). EGE (2), 152–155 (2008)
10. Prebeg, F.: Ekološko gorivo budućnosti (Ecological fuel of the future). Masmedia, Zagreb (2008)
11. Joksimović, V., Stevanović, M., Marjanović, Z.: Biofuels – Advantages and deficiency of use, Festival Kvaliteta 2008. In: 3rd Conference on the Quality of Life, Kragujevac, pp. 76–81 (2008)
12. US Department of Energy (April 10, 2010), `http://www.eere.energy.gov`
13. Biomasa kao obnovljivi izvor energije (Biomass as renewable energy source), Croatian Ministry of Agriculture, Forestry and Water Management, Croatian Ministry of the Economy, Labour and Entrepreneurship (April 10, 2010), Zagreb, more at `http://sumass.hr/assets/files/brosure/Biomasa.pdf`
14. SUGRE (Sustainable Green Fleets) (April 06, 2010), `http://www.sugre.info`

Basis of the Formalization and the Algorithmisation of the Control Functions in ATC Systems

Mariusz Maciejewski and Wiesław Zabłocki

Warsaw University of Technology, Faculty of Transport,
Koszykowa 75, 00-662 Warszawa, Poland
mariuszadammaciejewski@poczta.onet.pl, zab@it.pw.edu.pl

Abstract. The publication presents elements of the method of designing computer systems ATC, which consider the description of the control functions and also the interlocking functions on the basis of equations of the state describing the ATC system as a switching machine. The analysis of the ATC system as the switching machine allows distinguishing in the structure of this system a series of automatic components reflecting the courses and controlled objects. Equations of the state were used to describe these machines. This method enabled to formulate the interlocking function and equations, which may be applied to the algorithmization purposes. Moreover, the application of this method allows examining properties of these automata and hence of ATC system. The required properties of these automata include among others their controllability and observability.

Keywords: System, control, rail traffic, switching machine, automata, equations of the state, interlocking relations, functions.

1 Introduction

The assumption for formalization of the ATC system description is aimed at functions and interlocking equations, and as a consequence, at the reflection of the control task by the ATC system. The ATC system, which hereinafter is marked as $\mathbf{S}_{SRK}$ can be defined as an ordered triple (1) [9], [10], [11]:

$$S_{SRK} = (\boldsymbol{B}, S_S, U_S) \tag{1}$$

where: $\boldsymbol{B}$ – is a set of objects (devices), S_s – is a control system, and $\boldsymbol{U}_s$ – is a collection of control operations.

Because of the significance of the functions and tasks performed in the ATC system the object of the analysis is the S_S control system, because it plays the role of a central unit managing the whole system and the objects. The model of the S_S control system can be presented as follows [9], [11] and [12]:

$$MS_S = \{M_{STA}, M_{DYN}\} \tag{2}$$

where: M_{STA} – is a static model, and M_{DYN} – is a dynamic model.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 253–262, 2010.

The static model M_{STA} contains constant information on elements such as the control operations, the routing processes and the objects belonging to them, as well as on the required static properties taking part in the individual routes.

The dynamic model M_{DYN}, as having an importance for the considerations put forward in this publication, can be presented as follows:

$$M_{DYN} = (\boldsymbol{B}, \boldsymbol{U}_S, \boldsymbol{P}, \boldsymbol{AT}, \boldsymbol{T}_{DYN}, \boldsymbol{R}_{DYN}) \tag{3}$$

where: $\boldsymbol{B}$ – is a set of objects, $\boldsymbol{U}_S$ – is a collection of control operations, $\boldsymbol{P}$ – is a set of routing processes, $\boldsymbol{AT}$ – is a set of automata, which contains the automata subjected to the routing processes from set $\boldsymbol{P}$ and from set $\boldsymbol{B}$, as well as the input and output automaton, $\boldsymbol{T}_{DYN}$ – is a set of dynamic attributes, and $\boldsymbol{R}_{DYN}$ – is a set of relations established on the dynamic attributes.

The dynamic model M_{DYN} contains information on variable attributes of control operations and objects, as well as the set of relations describing connections occurring between them in time. These relations are constant for the given station and determine variable attribute information about:

- state of the system,
- routing processes which are currently held on the post,
- states of the objects.

The basis for the M_{DYN} model development is the set of system properties. The $M_{DYN}(MS_S)$ model should reflect these properties through reflecting the system states and their changes. Identification of the system takes into consideration the basic properties of the system:

- The ATC system can be reflected as an input – output model described as a certain switching machine,
- The switching machine reflecting the ATC system is a sequential machine, in which the output state depends on the current input states, and also on the sequence of the vectors of these signals,
- The system information can be reflected by discrete signals of binary values,
- The state of the system at a given moment of time is a set of all signal values occurring in the system,
- The sets of states, signals, and information of the ATC system are finite.

2 States of the ATC System

In the analysis of the system, which is a dynamic one, the concept of the system state, the state variable and the state space will be used. One distinguishes the system states describing the control system and the objects. The basic system state is a fundamental state. The set of all possible system states, i.e. of the control system and of the objects, determines the space of system states. The change of any unit of the system (i.e. the change of any signal of this unit) means the change of the state of the whole system. The system states can be presented in the form of simplified graph of states, which is depicted in Fig. 1. A change of the control system state causes a change of the objects' state, and a change of the objects' state means also a change of the system state.

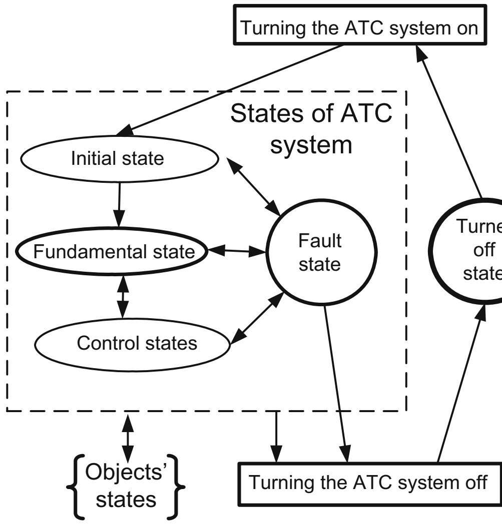


Fig. 1. Distinguished states of the ATC system. The distinguished states are defined by state variables, which describe the functions, including the interlocking functions. The system can be in the following states: turned off, initial, fundamental, performing the control task and fault.

Turned off state
This is the state when the system does not perform any function. This state is identical to physical disconnection of the voltage supplying the computers or this state occurs while turning on the computer, running the operating system and software application.

Initial state
This is the state when all software applications run and all functions are performed. During this state the system assigns the initial values to the state variables, which describe the objects (elements and devices) according to the specification prepared by the system designer.

Fundamental state
The fundamental state of the system is such a state, when no running is carried out and external actuators are in the fundamental states appropriate for them, i.e. in the stand-by state defined by the designer. The specific feature of the ATC system is that the system returns to the fundamental state after releasing all routes (after completion of the control task), and when all objects will return to the fundamental state and will report readiness to work.

Control state
The control state is a state when the control tasks are being carried out, initiated by the commands to perform routing process or by individual commands. This is the state, in which even only one route is in a state different from the fundamental state, and the objects are in good working order and are ready to work, but their state variables representing the internal state reflecting the state of the routing process, and devices differ from the fundamental state of the objects.

Fault state
This is the state when the object has a limited controllability. The state variables of such object show the malfunctioning of the computer system components and/or the failure of the device connected with the object.

3 Control System as an Automaton

The above mentioned properties of the ATC system thus allow for identification of the S_S system as a certain automaton, which may be a classic automaton: abstract, sequential, discrete, finite or deterministic one, because relations between input and output signals are unique and unchangeable. The automaton having the above mentioned properties is presented in the professional literature as so called ordered "quintuple" [5]:

$$A = (\mathbf{X}, \mathbf{Y}, \mathbf{Q}, \delta(X, Q), \lambda(X, Q)) \tag{4}$$

where: **X** – is a set of input states of X^i, **Y** – is a set of output states of Y^i, **Q** – is a set of internal states (memory states) of Q^i, δ – is a transition function, and λ – is an outputs' function.

The individual states of X^i, Y^i and Q^i are represented by the signal vectors, ***X, Y*** and **Q,** respectively (6). The operation of the automaton is described by equations (6): of the state and the output, respectively:

$$X = (x_1, x_2, ..., x_r),\ \ Y = (y_1, y_2, ..., y_m)\ and\ Q = (q_1, q_2, ..., q_3) \tag{5}$$

$$\begin{aligned} Q^{t+\tau} &= \delta(Q^t, X^t) \\ Y^{t+\Delta} &= \lambda(Q^t, X^t) \end{aligned} \tag{6}$$

where: τ – time of processing delay, after which the new state of $Q^{t+\tau}$ is determined to differentiate from the previous state of Q^t,

Δ – time of processing delay, after which the new state of $Y^{t+\Delta}$ is determined to differentiate from the previous state of Y^t,

δ – transition function $\delta(Q^t, X^t)$, which determines the changes of automaton's memory states,

λ – outputs' function $\lambda(Q^t, X^t)$, which determines the changes of outputs' states.

The dynamic model reflected by means of the automaton has the following properties [10], [11], [12]:

- the state of X^i inputs (***X*** signals vector) depends on tasks coming from the system environment, on the messages from the objects and on the system owned events,
- the state of Y^i outputs (***Y*** signals vector), which comprises commands to the objects and information transferred into the system environment,
- Q^i memory state – is a ***Q*** vector representing internal state (system memory state),
- the transition function and the outputs' function reflecting the control relations comprising among other things an operator's functions, interlocking functions and other functions.

Performing the transformation of the general form of equations (6) one obtains the set of equations (7) which determine the values of individual state variables q_i and output signals y_i. Equations (7) describing the state variable signals and output signals show the principle for creation of the control function and equations.

$$
\begin{aligned}
q_1^{t+\tau_1} &= \lambda_1(q_1^t, q_2^t, \ldots, q_n^t, x_1^t, x_2^t, \ldots, x_r^t) \\
q_2^{t+\tau_2} &= \lambda_2(q_1^t, q_2^t, \ldots, q_n^t, x_1^t, x_2^t, \ldots, x_r^t) \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
q_n^{t+\tau_n} &= \lambda_n(q_1^t, q_2^t, \ldots, q_n^t, x_1^t, x_2^t, \ldots, x_r^t) \\
\\
y_1^{t+\Delta_1} &= \delta_1(q_1^t, q_2^t, \ldots, q_n^t, x_1^t, x_2^t, \ldots, x_r^t) \\
y_2^{t+\Delta_2} &= \delta_2(q_1^t, q_2^t, \ldots, q_n^t, x_1^t, x_2^t, \ldots, x_r^t) \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
y_m^{t+\Delta_r} &= \delta_r(q_1^t, q_2^t, \ldots, q_n^t, x_1^t, x_2^t, \ldots, x_r^t)
\end{aligned}
\qquad (7)
$$

The graphical interpretation of equations (7) is depicted in Fig. 2.

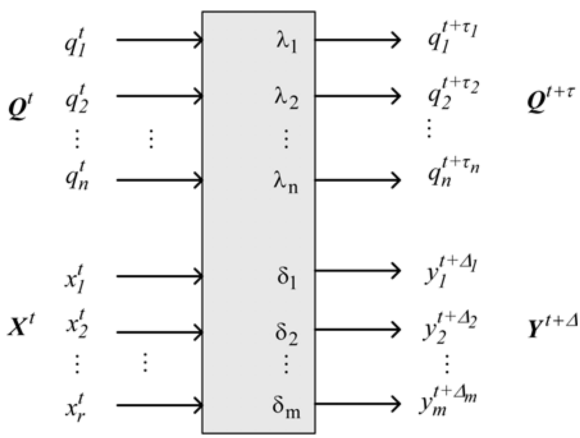


Fig. 2. The principle for creation of a new vector of memory state and outputs' state

The rectangle symbol represents the set of functions, which process the memory state signals present at the time of *t* and input signals by calculating the new signals of internal state and output signals. The above mentioned functions do not indicate in which sequence the individual memory state variables will be calculated. This is essential, because the state variables are mutually dependent and also they are dependent on the remaining state variables, and in turn the output signals depend on the state variables. The solution of this problem comes down to calculation of the memory state signals and output signals in the subsequent discrete steps in time – see Fig. 3. In each step at the beginning of the cycle a one-time reading of the current vector of the memory state signals and the vector of input signals takes place, on the basis of which the δ and λ functions determine the current values of the memory state vector signals and the vector of output signals.

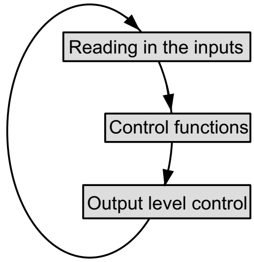


Fig. 3. The principle of sequentiality in the information processing. One working cycle comprises the following activities: reading in the inputs, information processing and output level control.

In the computer based ATC systems one assumes the constancy of the processing cycle, whereas the cycle length must be selected so that all changes of the input signals could be observed – and the duration of cycle change must be short enough to ensure that each event of the input signals change could be registered. The advantage of sequential signal processing is the elimination of hazard phenomena.

4 Analysis of Control System Automaton

Presented description of the properties of the A automaton, which would reflect the changes of state of the S_S control system can be used to describe the AT automaton. In the case of automaton A the automaton structure is unknown and only the automaton definition is known. In the case of automaton AT it is possible to suggest a certain structure of the automaton. Assuming that automata A and AT will be equivalent one another (8) further considerations can be conducted in relation to automaton AT.

$$A \equiv AT \tag{8}$$

Let the analysis of the tasks related to the routing processes, as well as to the individual objects, i.e. the outside devices, be the basis for determination of automaton AT structure. Taking into consideration the concept of the routing process state established in this publication one can create the set of states of each routing process. The states of the routing process as well as the changes of these states can be reflected by means of automata. The set of automata assigned to individual routing processes is the ***AP*** set. In a similar way one determines the set of automata ***AB*** containing the states as well as the changes of the states of individual objects and thus subjected to these objects. Due to the necessity of handling physical input and output signals and adaptation of these signals to the processing requirements a single automaton of input – *Awe* and output – *Awy* have been introduced. By means of these automata it is possible to perform relations with the environment and the outside devices. Because of this the AT automaton is a sum of automata ***AT***: *Awe* and *Awy*, the sets of automata ***AP*** and automata ***AB*** (9).

$$\boldsymbol{AT} = Awe \cup \boldsymbol{AP} \cup \boldsymbol{AB} \cup Awy \tag{9}$$

The commands from the collection of control operations are introduced into the ***AT*** automata. Other commands are sent to the ***B*** object set. Information on the system state is created on the basis of the messages from set ***B*** and on the basis of signals of memory state and automata outputs and analogous object signals. On the basis of automaton ***AT*** structure one creates the data structures of fixed (static model) and variable attributes describing the S_S control system and objects. The vector signals of the memory state and input vectors of all automata belonging to the ***AT***, i.e. to the routing processes and objects, are processed at subsequent moments after previously reading in the signals of input vectors. While performing the identification of individual automata types one determines the group of signals according to the input – output model and the properties of sequential automata. The separate groups of signals are well-ordered and arranged in the forms of vectors.

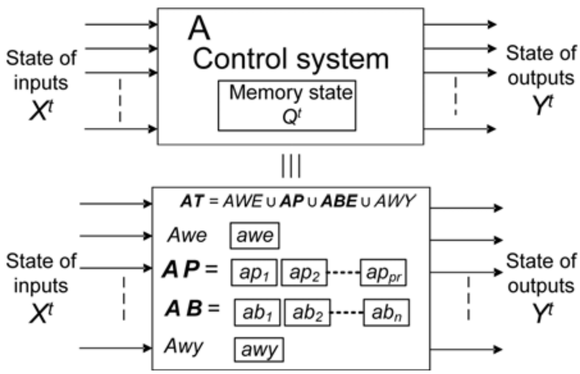


Fig. 4. The control system (controller) structure as a structure composed of multiple automata

On the basis of a functional description of an automaton one selects the state variables for this automaton and designs the transition functions describing the individual selected state variables and transition conditions between the state variables. The transition conditions for the whole system require elaboration of the graph on which one indicates the individual state variables (nodes) and arcs described with the transition functions. In this way one acts in relation to each type of automata. The selected issues concerning the process of automaton designing will be presented on the example of route automaton.

5 Route Automaton

The route is a dynamic category, because during the train running the indication of signalling devices and the states of railway section controlled and registered by the system are changing. The changes of signalling devices indications and the sequence of changes of the railway section states taking place during the train running create the condition for the route to be released. The state of routing process is also reflected

Table 1. The list of transition functions λ_p of the route automaton

	Functions λ_p and FPSZUKP(), FPSZUKM(), FPSPRP(), FPSPRM()		
1.	$\lambda_{p(br,szukp)} = q_{pbr} \bullet r_{pup}$	8.	$\lambda_{p(sprm,br)} = q_{psprm} \bullet$! FPSPRM()
2.	$\lambda_{p(br,szukm)} = q_{pbr} \bullet r_{pum}$	9.	$\lambda_{p(sprp,utw)} = q_{psprp} \bullet$ FPSPRP()
3.	$\lambda_{p(szukp,br)} = q_{pszukp} \bullet$! FPSZUKP()	10.	$\lambda_{p(sprm,utw)} = q_{psprm} \bullet$ FPSPRM()
4.	$\lambda_{p(szukm,br)} = q_{pszukm} \bullet$! FPSZUKM()	11.	$\lambda_{p(utw,zwol)} = q_{putw} \bullet$ variable indicating the confirmation of all sections of the course of route
5.	$\lambda_{p(szukp,sprp)} = q_{pszukp} \bullet$ FPSZUKP()		
6.	$\lambda_{p(szukm,sprm)} = q_{pszukm} \bullet$ FPSZUKM()	12.	$\lambda_{p(zwol,br)} = q_{pzwol} \bullet$ variable indicating the confirmation of all sections of the course of route
7.	$\lambda_{p(sprp,br)} = q_{psprp} \bullet$! FPSPRP()		
	FPSZUKP, FPSZUKM – these functions take the value of "1" when the train route or manoeuvring route will be found, respectively.		
	FPSPRP, FPSPRM – these functions take the value of "1" when the condition for confirmation of the train route or manoeuvring route are met, respectively.		

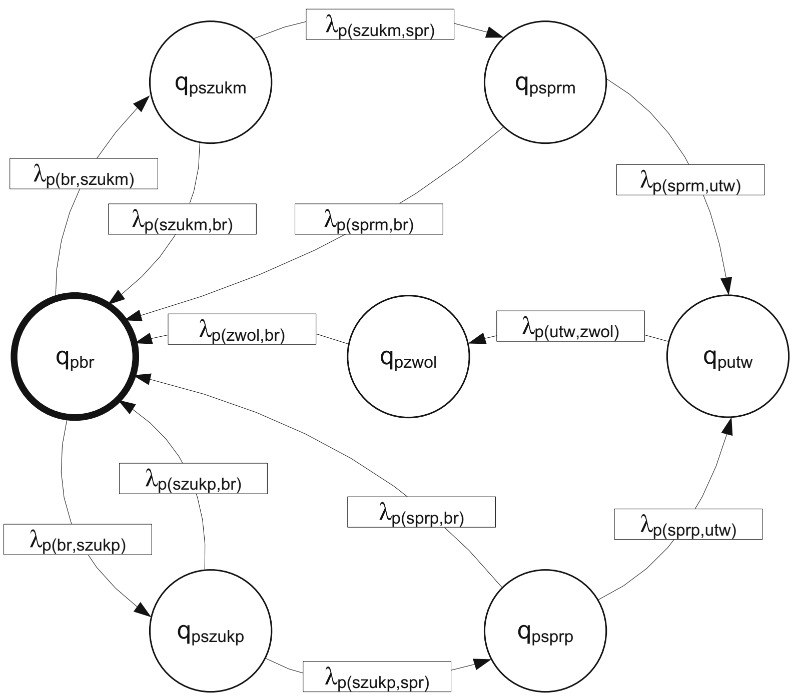


Fig. 5. Graph of route automaton (ap)

by the states of the objects belonging to the course of a given route. In the route automaton seven variable states and twelve functions have been distinguished, on the basis of which the values of these variables are determined. The signals of state variables which create at the same time the state vector of internal routing process are as follows:

1. q_{pbr} – the state when there is a lack of confirmed route (non-confirmed route),
2. q_{pszukp} – the state of searching for the train route,
3. q_{pszukm} – the state of searching for the manoeuvring route,
4. q_{psprp} – the state of validation of the train route,
5. q_{psprp} – the state of validation of the manoeuvring route,
6. q_{putw} – the state of confirming elements in the course of route,
7. q_{pzwol} – the state of awaiting for route to be released.

The transition functions λ_p for the route automaton are listed in Table 1, and their graph is depicted in Fig. 5.

6 Conclusions

The issues presented in the publication, however related to the classic method of the ATC system description on the basis of the state equations, are embraced by the method of computer system design. This method takes into consideration the subsequent systematic design phases, starting from the informal description of the ATC system, through the system assumptions, formal description of the individual automata creating the control system, description of control functions, design of software and hardware configuration, and system implementation and installation.

Thanks to the ATC system design method a number of systems in railways and underground stations have been developed and set working, of course after obtaining the appropriate certificates.

A successful implementation of the method does not close it. It is expected that research extending the automation and designing will be continued, and that the method which allows determining the controllability and observability of the individual automata types will be developed.

References

1. Bushkov, V., Yevtushenko, N., Tiziano, V.: Discussion on Supervisory Control by Solving Automata Equation. In: Proceedings of the IEEE, EWDTS 2009, pp. 77–80 (2009)
2. Findeisen, W. (ed.): A joint publication: System Analysis – Basis and Methodology Analiza systemowa – podstawy i metodologia. Wydawnictwo Naukowe PWN, Warszawa (1985) (in Polish)
3. Kaczorek, T.: Theory of Control and Systems (in Polish) Teoria sterowania i systemó. Wydawnictwo Naukowe PWN, Warszawa (1999)
4. Lin, F.: Robust and Adaptive Supervisory Control of Discrete Event Systems. IEEE Transaction on Automatic Control 38(12) (December 1993)
5. Traczyk, W.: Digital Circuits. Theoretical Basis and Methods of Synthesis. In: Układy cyfrowe. Podstawy teoretyczne i metody syntezy. WNT, Warszawa (1984) (in Polish)
6. van Vlijmen, S.F.M.: Verification of the Vital Processor Interlocking. In: FMERail Workshop 1, Netherland (1996)
7. Wanga, W., Lafortunea, S., LinbAn, F.: An algorithm for calculating indistinguishable states and clusters in finite-state automata with partially observable transitions. Systems & Control Letters 56, 656–661 (2007)

8. Yang, Z., Blanke, M.: A unified approach for controllability analysis of hybrid control systems (2000), http://www.iau.dtu.dk/secretary/pdf/yang_blanke_2000b.pdf
9. Zabłocki, W.: A Formal Analysis of Conflict Functions Used in Rail Traffic Control Systems. Archives of Transport, Polish Academy of Sciences 18(3), 81–99 (2006)
10. Zabłocki, W.: Interlocking Functions of ATC Station System. Archives of Transport Polish Academy of Sciences 4, 89–108 (2008) ISSN 0866-9546
11. Zabłocki, W.: Modelling of Automatic Train Control (ATC) station system (in Polish) Modelowanie stacyjnych systemów sterowania ruchem kolejowym. Oficyna Wydawnicza PW, Warszawa (2008)
12. Zabłocki, W.: Selected problems of the description of interlocking function in ATC systems. In: 17th International Symposium EURO Żel 2009, Increasing the competitiveness of the European Rail System, pp. 207–216. Univerzita v Žiline, Žilinska (2009)

An Application of Traffic Measurements to Route Planning for Traffic Flow Simulation in MATSim

Michał Maciejewski

Institute of Machines and Motor Vehicles, Poznan University of Technology,
ul. Piotrowo 3, 60-965 Poznan, Poland
michal.maciejewski@put.poznan.pl

Abstract. The paper presents MATSim Junction Turning Ratios Route Planner (JTRRouter) that enables to generate route plans for vehicles on the basis of traffic measurements conducted at intersections. The authors described a general design, functionality, and requirements of JTRRouter, and then presented a sketch of the route planning algorithm and discussed its computational complexity. In order to illustrate the most essential capabilities of the module, an example of route planning for a fragment of a real urban network was presented.

Keywords: Traffic flow, modeling, simulation, traffic measurements, route planning, MATSim, TRANSIMS, SUMO.

1 Introduction

In order to obtain simulation results that are reliable and consistent with the real traffic, besides a precise network model, a correct traffic generation method is required. Such a method may be based on data derived from traffic flow measurements performed at intersections.

Unfortunately, despite great functionality and versatility of MATSim, it does not contain a tool that performs route planning on the basis of intersection turning ratios. To overcome this limitation a special module, MATSim JTRRouter (MATSim Junction Turning Ratios Router), was implemented in Java. The paper presents more detailed information about assumptions, constraints, and functionality of the module. Furthermore, the input data formats are presented and light is shed on the details of the implemented route planning algorithm.

2 MATSim

MATSim (Multi-Agent Transport Simulation) [1], [2] is an open source multi-agent activity based microsimulation system for a daily transport demand analysis. Due to fast and efficient traffic simulation, it enables to conduct the analysis for large scenarios, even concerning the whole country. MATSim consists of several modules, each one responsible for certain tasks within the complete process of transportation modeling and simulation. As it is open source, one can easily extend MATSim functionality and adjust it to one's needs.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 263–270, 2010.

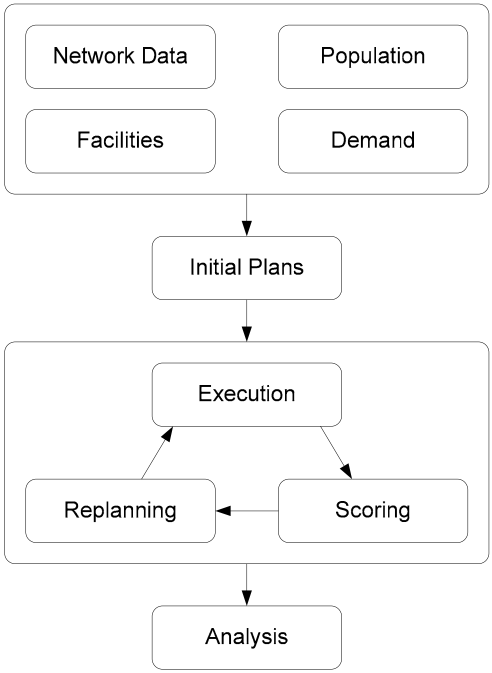


Fig. 1. Standard modeling and simulation process flow in MATSim

A schema of a typical modeling and simulation process flow in MATSim is presented in Fig.1.

After the network and facility databases are prepared, a synthetic population is generated, and then initial demand modeling is performed. As a result, each individual in the population has his/her own initial daily activity plan (daily plans consist of work, shopping, school, and other kinds of activities). Afterwards there are three steps (execution, scoring, and replanning) that are usually run iteratively in order to obtain system equilibrium according to the first Wardrop's principle [3]. The plans execution consists in queue-based traffic simulation, the scoring assesses the execution of daily plans for each individual, whereas the replanning step optimizes daily plans for some selected individuals. After the process of the iterative demand optimization is finished, the obtained results are analyzed in detail.

3 MATSim JTRRouter Module

3.1 Idea

Unfortunately, the original process flow is not always desirable or even feasible. When detailed measurements of traffic flow are available, the route planning procedure can use the measured data to construct routes that emulate the real traffic. In such cases, one may even skip the process of the iterative demand optimization, as well as the initial demand generation. In consequence, the process flow may, in the simplest case, be non-iterative and very straightforward, as shown in Fig.2.

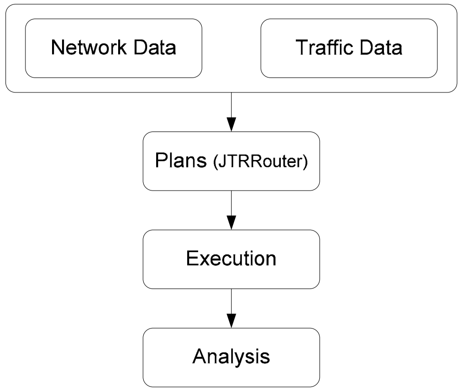


Fig. 2. Simplified modeling and simulation process flow in MATSim

In this paper, the authors consider an application of traffic volume counts performed at intersections to vehicle routes planning. A similar approach is implemented in JTRROUTER program, which is a part of SUMO microsimulation system [4]. Therefore, the route planner module implemented by the authors was called MATSim Junction Turning Ratios Router, or simply JTRRouter, after its equivalent in SUMO. The authors had previously implemented an analogous module for TRANSIMS microsimulator [5], [6].

MATSim JTRRouter was created for and tested with MATSim version 0.1.1.

3.2 Network Model

An application of JTRRouter imposes some constraints on a network model. To illustrate them, let us consider a network presented in Fig.3. The network is built of nodes (presented with identifiers) and one-way links. Nodes 1,…,6 are intersections, and 10,…,18 are boundary nodes, where each route starts and ends. Since all roads are two-way, they were modeled as pairs of one-way links of opposite directions. All left turns at intersection 6, and two left turns at intersection 4 (6→4→3 and 2→4→14) are prohibited. The network model represents a fragment of a real road network in Poznan, Poland, and was used in simulation research [7], [8].

3.3 Input and Output Files

JTRRouter requires two input XML files (flows.xml and turns.xml) containing data derived from the traffic measurements, and produces one output XML file (plans.xml) necessary to carry out the simulation. JTRRouter also requires a standard MATSim file with a network definition (network.xml).

The first input file, flows.xml, contains flow definitions specified for each of the boundary nodes (nodes 10-18 in the example). Below there is presented a fragment of flows.xml file. The first line defines simulation period in seconds (*startTime* and *stopTime*) and a flow coefficient (*flowCoeff*) used for a proportional increase/decrease of all flow rates. Line 2 specifies flow rate for boundary node 11. According to this line, link 101 (*inLink*; 11→1) is a source for an incoming flow of vehicles, and link 201 (*outLink*; 1→11), of the opposite direction, is a sink for an outgoing flow of vehicles. The flow rate of the incoming flow is defined as 801 vehicles per hour (*no*). If *inLink* or *outLink* attribute is omitted, then, at the given boundary node, there is no incoming or outgoing flow of vehicles, respectively.

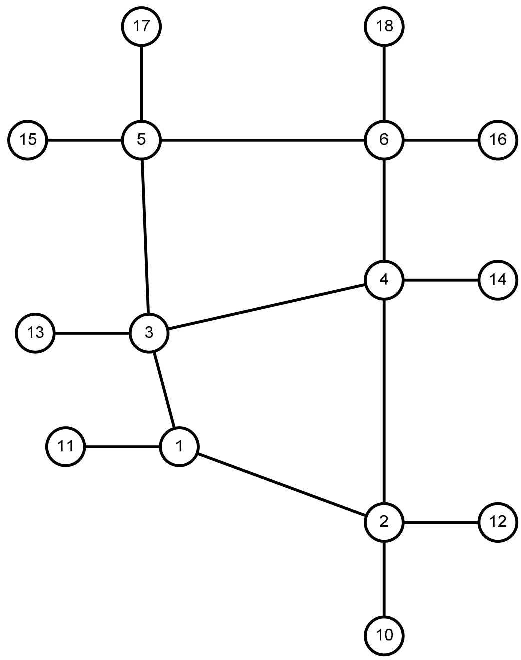


Fig. 3. Model of the road network

Fragment of flows.xml file containing flow definitions

```
<flows startTime="0" stopTime="7200" flowFactor="2.0">
  <flow inLink="101" outLink="201" no="801"/>

  <!-- other "flow" entities -->

</flows>
```

Traversing through the network, at each intersection inlet, a vehicle/driver selects one of all allowed maneuvers. The selection may be stochastic or deterministic, and depends on the measured turning ratios. The turning ratios for each intersection inlet are specified in the second input file, turns.xml, where, for each inlet, a set of possible outlets (i.e. next nodes) with their selection probability is defined.

Fig.4 illustrates the possible turning maneuvers for vehicles approaching the intersection 1 from boundary node 11. Each driver selects the next node (node 2 or 3) according to the defined ratios. In a snippet of turns.xml file presented below, the ratios were defined as 0.671 and 0.329 for choosing node 2 or 3, respectively.

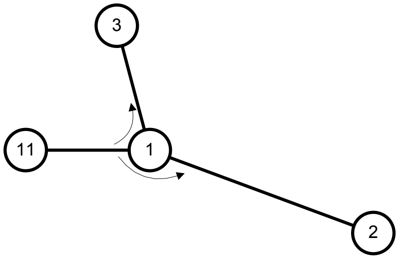


Fig. 4. Possible turning maneuvers at intersection 1 for vehicles approaching from node 11

Fragment of turns.xml file containing turning ratios definitions for each intersection inlet

```
<turns>
  <turn curr="1" prev="11">
    <next id="2" probability="0.671"/>
    <next id="3" probability="0.329"/>
  </turn>

  <!-- other "turn" entities -->

</turns>
```

The output of JTRRouter plans.xml file contains a detailed information about all the planned routes (departure nodes, departure times, lists of visited nodes, etc.). This file and network.xml file, which defines the network model, are the input for the simulation (execution) module.

4 MATSim JTRRouter Algorithm

The general scheme of the route planner algorithm is presented below. The main part of the algorithm (lines 5-17) consists in finding all acceptable routes, and assigning them (with the previously calculated selection probabilities) to each generated vehicle.

Sketch of JTRRouter algorithm

```
1.  READ startTime, stopTime, flowCoeff
2.  READ flows
3.  READ turns
4.  READ network

5.  INIT plans

6.  FOR EACH f IN flows
7.      routes = findRoutes(f, network)

8.      FOR EACH r IN routes
9.          r.probability = calcSelectProbability(r)
10.     END FOR EACH

11.     flowPlans = generatePlans(f)
```

```
12.     FOR EACH p IN flowPlans
13.         p.route = selectRoute(routes)
14.         p.time = calcDepartureTime(r, f)
15.         add(plans, p)
16.     END FOR EACH
17. END FOR EACH

18. WRITE plans
```

The procedure `findRoutes` (line 7) finds all acceptable routes starting at a given boundary node (i.e. for a given incoming flow *f*). A route is acceptable if no link is traversed twice (links are one-way). On the other hand, visiting any node twice is allowed. The procedure was implemented as a depth-first search that is run recursively until a boundary node is reached, where a vehicle exits the network, or any link is traversed for the second time, which is forbidden.

A roughly approximated upper bound of computational complexity for `findRoutes` procedure is $O(bn^m)$, where:

- b – number of boundary nodes
- n – number of intersection nodes
- m – number of (one-way) links between nodes.

This is a very high complexity, but in the case of typical road networks, which are sparse graphs, the upper bound is significantly lower. For a typical road network, with three allowed maneuvers (right turn, straight ahead, left turn) at each intersection, a very rough upper bound is $O(b3^m)$. The actual computational complexity is, however, much lower, because the recursion search rarely goes deep, since most of the search branches are cut on shallow recursion levels. The cuts are made due to reaching one of the boundary nodes, or, which is mostly the case, traversing any link for the second time. Moreover, in reality a lot of turning maneuvers are prohibited (e.g. some left turns, turning into outlets of one-way streets), which also shortens computation time.

Since `findRoutes` procedure is called iteratively for each boundary node (i.e. incoming flow *f*), the approximated upper bound of computational complexity for the whole algorithm is roughly $O(b^2n^m)$.

A route selection probability *r.probability* (line 9) is equal to the product of probabilities of all turning maneuvers within the route. To improve the computational efficiency, the route selection probabilities can be computed within `findRoutes` procedure.

All plans, for a given incoming flow *f*, are generated according to the measured traffic volume (line 11). Then, a route and a departure time are assigned to each plan. However, the assignment procedure (both the route selection and the time determination) may be deterministic or stochastic (lines 12-16).

Fig.5 shows an example of four routes generated by the route planner algorithm for the network presented in Fig.3. Despite the high upper bound of computational complexity, the procedure runs fast for typical road networks. In the case of the presented network, it took only a few milliseconds to find all 308 acceptable routes and generate plans for over 6,300 vehicles (excluding time of I/O operations).

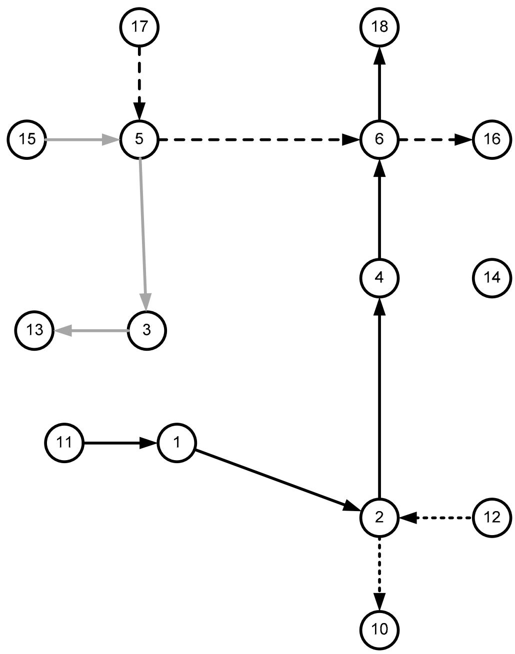


Fig. 5. Example of routes generated by JTRRouter

5 Conclusions

The goal of JTRRouter is to enable to use traffic measurements for the initial plans generation process. JTRRouter, in both MATSim and TRANSIMS versions, was successfully applied by the authors in several research projects [7], [8]. As it was shown, the implemented algorithm provides fast route plans generation. In the future, it is planned to enable to define default turning ratios (for intersections without turning ratios explicitly specified), and to facilitate minimization of discrepancies in the traffic measurement data.

References

1. Balmer, M., Meister, K., Rieser, M., Nagel, K., Axhausen, K.W.: Agent-based simulation of travel demand: Structure and computational performance of MATSim-T. In: 2nd TRB Conference on Innovations in Travel Modeling, Portland (2008)
2. Multi-Agent Transport Simulation, MATSim, `http://matsim.org/`
3. Wardrop, J.G.: Some theoretical aspects of Road traffic research. Proceedings of the Institution of Civil Engineers 1(2) (1952)
4. Simulation of Urban MObility, SUMO, `http://sumo.sourceforge.net/`

5. Maciejewski, M.: Junction turning ratios route planner module for TRANSIMS. In: IInd International Scientific Conference “Transport Problems”, Katowice-Kraków, Poland, CD (2010)
6. TRansportation ANalysis and SIMulation System, TRANSIMS, `http://www.transims-opensource.net/`
7. Maciejewski, M.: A comparison of microscopic traffic flow simulation systems for an urban area. In: IInd International Scientific Conference “Transport Problems”, Katowice-Kraków, Poland, CD (2010)
8. Maciejewski, M.: Kalibracja parametryczna kolejkowego modelu przepływu ruchu drogowego w systemie MATSim. In: Proceedings of VII Konferencja Naukowo-Techniczna “Systemy transportowe. Teoria i praktyka”. Katowice, Poland (in print, 2010)

Public Transport Information System for Visually Impaired and Blind People

Michał Markiewicz and Marek Skomorowski

Institute of Computer Science, Jagiellonian University, ul. Prof. Łojasiewicza 6, 30-348 Kraków, Poland
{Michal.Markiewicz,Marek.Skomorowski}@ii.uj.edu.pl

Abstract. This paper presents an assistive system for the visually impaired and blind people which helps them using public transport means. The proposed system uses mobile phones as a medium for passenger information system and GPS (Global Positioning System), GSM (Global System for Mobile Communications) and Bluetooth technologies for location and communication purposes. In the proposed system sound messages are given to the blind people via mobile phones which have dedicated software installed. This system has been implemented and tested in public transport in two pilot cities.

Keywords: Visually impaired, blind people, public transport, mobile phones.

1 Introduction

The visually impaired and blind people encounter difficulties in navigating independently in urban areas. Therefore, the use of an assistive mobile system seems to be necessary in the case of the visually impaired and blind people. The engineering and design principles and techniques used in the assistive technology for blind and visually impaired people are presented, for example in [1], [2], and [3]. However, the existing assistive systems for mobility of the visually impaired and blind people in public transport in cities are not satisfactory, which was the motivation for continuing research on the subject.

This paper presents an assistive system, called *Ariadna* for improving mobility of the visually impaired and blind people in public transport. The goal behind creating Ariadna was to develop a mobile passenger information system, which provides real-time information in public transport means for visually impaired and blind people. The proposed system has been implemented and tested in pilot cities in Poland, namely in Warsaw and Nowy Sącz.

2 Types of Barriers to Visually Impaired and Blind People in Public Transport

Public transport is vital to visually impaired and blind people. Local governments and non-governmental organizations are interested in investigating difficulties in using

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 271–277, 2010.

public transport means. They are creating reports, which helps improving public transportation and making it usable for people with disabilities. An example of such a report can be found in [4].

The main problems described in [4] and [5] are:

1. Lack of vehicle identification, which means that a blind person is not properly informed, that the vehicle that have just arrived is the member of the public transport fleet (in the case, that vehicles which belong to other transport companies can share the stop).
2. Stops are not properly marked; there is no possibility to instruct a blind person where the bus stop is.
3. Lack of acoustic information about line numbers of arriving vehicles.
4. Lack of timetable in format which is readable by the blind people.
5. Lack of possibility to inform the driver to take care about a blind person, who is waiting at the stop.

The most of these barriers are easily worked around by legal regulations, which allow traveling for free for the people that take care about blind people and assist them during their voyages. To really solve the problem and allow blind people to travel independently without assistance there is a need of an information system, which would give them up-to-the-second notifications of arriving vehicles, access to the timetable in a way that can be accessed at the stop and moreover allow the driver to be notified of a disabled passenger waiting.

Many cities introduce their own solutions, which partially solve the problems mentioned. They will be briefly described in the next chapter.

3 The Selected Assistive Systems for the Visually Impaired and Blind People Used in Public Transport

One of the best known assistive mobile systems for the visually impaired and blind people is the system in Vienna in Austria. The idea of the system is the following: the visually impaired and blind people have electronic remotes. The electronic remotes are turned on by the visually impaired and blind people when they hear that a vehicle is arriving. Each vehicle is equipped with an electronic device which reacts to the electronic remote signal and emits the number of the line and the direction of the route. This system has been deployed in some cities in Poland like Bydgoszcz, Łódź.

In Prague in the Czech Republic all vehicles are equipped with a special electronic device, informing the visually impaired and blind people about the route number and the destination of the approaching vehicles. The same electronic device gives the driver acoustic information of the intention of impaired and blind people to get on the vehicle. Automatic announcement of stops onboard vehicles helps to orientate along the route.

To give the information about the city topography the cities introduce various solutions. For example in the New York City it is possible to get information about city subways to riders who are blind or visually impaired using a booklet of raised-line and textured maps of train routes [6].

There are some other solutions addressed to the people with different disabilities like the project ASK-IT for the mobile impaired people, which helps them in planning the trips using accessible means of transport [7].

To summarize, the existing solutions partially remove the barriers mentioned before. Usually they require proprietary equipment, which is not in mass production. This often results in a high price of deployment of such systems in the cities. We considered it as the main disadvantage. Beside this, another problem that we wanted to resolve is the provision of a more holistic approach, which provides not only the information about arriving vehicles or information about the location of a stop, but at low cost eliminates all the barriers mentioned in the previous chapter and is dedicated to the visually impaired and blind people. The proposed system is described in the next chapter.

4 Description of the Proposed System

In order to enable visually impaired and blind people to travel freely using the public transport means we decided to focus on the following issues:

- To provide on time precise information about arriving vehicles,
- To give access to the timetable and current delays,
- To give access to the online route planner,
- To inform the driver about a disabled person waiting,
- To use standard devices which are in mass production, namely mobile phones as communication devices for the blind people and PDAs as the terminals for the bus drivers.

The outline of the proposed system is the following: a visually impaired or blind person turns on the application on the mobile phone, then checks out the code of the stop (which is written in Braille on the timetable) and types it on the keypad. It is also possible to automatically detect the number of the stop, provided that the mobile phone used by the blind person is equipped with a GPS receiver. In that case the geographical coordinates are compared with the predefined list of stop locations and based on that the nearest stop is selected. This allows also helping to locate that stop, if the blind person knows the stop he/she wants to visit. After that the blind person is subscribed to the distribution list of all events related to this stop, namely the person receives all the information about arriving vehicles, delays and notifications sent by the dispatcher to the awaiting passengers. In addition, the driver of the vehicle is informed about disabled people waiting at the stop (this additionally requires entering by the user the number of the line he is waiting for).

Opposite to the existing solutions, the proposed system uses a central server to mediate between the public transport vehicles and the users and does not rely on the proximity detection. This is because we used the most common communication devices in the world – mobile phones – which are typically not equipped with proximity sensors that could be useful for this purpose. Therefore, the user has to provide the information about his location. Stop codes are short (four digits) and unique, so it is easy to remember the codes of a few most visited stops (near home and office, for example).

Vehicle equipment, which allows the driver to receive notifications of awaiting disabled passengers, is a terminal with built-in GPS receiver and the module which provides wireless communication with the central server. The terminal is continuously sending the information about vehicle's position on the route to the central server. After the reception of the information, that the distance to the next stop on route is less than the predefined threshold, the information about arriving vehicle is relayed to all passengers waiting on the stop (and who gave this information to the system by typing stop code or by allowing the built-in GPS to send coordinates of their position). As a result passenger's mobile phone vibrates, and emits an appropriate communication messages (which can be different according to the personalized setting such as preferred language).

The hardware which is required to run the system is as follows: mobile phones for the end users (Java enabled, possibly with the GPS built-in), PDA's (or on-board computers which can wirelessly transmit vehicle's position to the system center and display notifications to the driver) and the main server which is responsible for handling the communication between these two groups of devices.

The architecture of the proposed system is presented in Fig. 1.

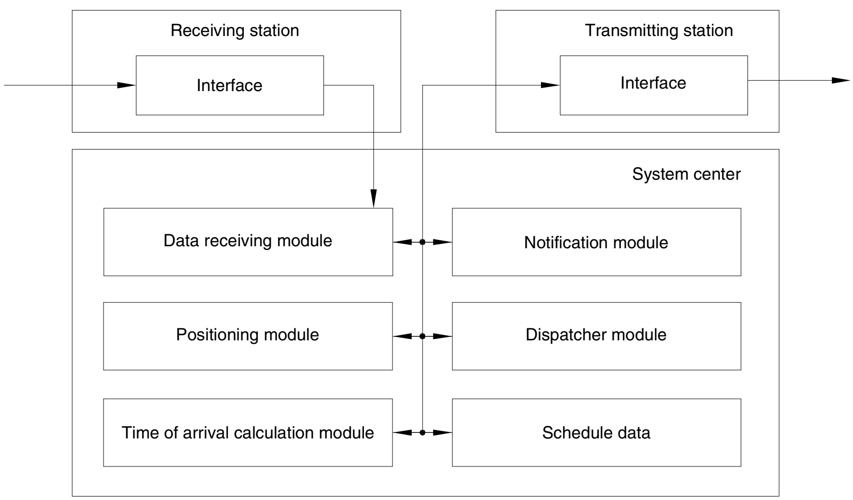


Fig. 1. The architecture of the system

The software part can be divided into three applications, which are running on mobile phones, PDAs and on the central server:

- The application for mobile phones, which has a form of MIDlet that after start up connects to the central server, and then plays sequences of messages notifying of events, for example "The bus number 139 is arriving."
- The application for drivers, which continuously sends the geographical coordinates of the vehicle to the central server and displays notifications to the driver.

- The central application responsible for routing messages coming from devices installed in the vehicles to the mobile devices owned by passengers.

The Application for mobile phones was designed in strict collaboration with the blind people. The main screen of the application consists of only a single text field for entering the code of a stop. The application was written in J2ME and has working-profile MIDP 1.0 and CLDC-1.1 configuration.

The application for PDAs (Personal Digital Assistants) was implemented in .Net Compact Framework. After the start up it connects with the central server, and starts sending the data about the current location of the vehicle. Moreover it allows the driver to enter the information about the current line number and displays the information about passengers waiting on the bus stops. The application is also able to play the bus stops announcements inside the vehicles.

The information about positions of public transport vehicles is sent to the central server using the General Packet Radio Service (GPRS) of the GSM network. The same protocol is used for the communication between mobile phones and the central server.

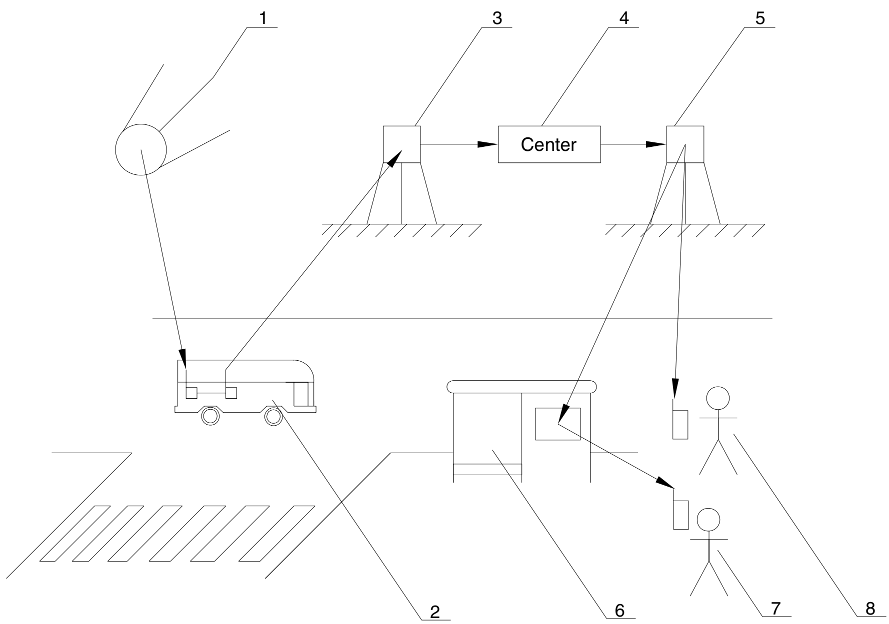


Fig. 2. The GPS (1) satellites provide the data required to get the information about the current position of a public transport vehicle (2). This information is sent to the receiving station (3) using the GPRS protocol and in the system center (4) it is used to determine when to send the information about a possible delay or notification of vehicles arriving by the transmitting station (5) using the GPRS protocol to the mobile phones owned by end users (8). The system extension allows sending the information to the module installed on a bus stop (6), which is then distributed to the end users (7) by a Bluetooth interface.

Although the main functionality is to provide the information about arrivals of vehicles, the proposed system allows:

- Checking the time of arrival of the line of the next arriving vehicle
- Checking the time of arrival of the next arriving vehicle of the given line number
- Getting the list of all line numbers, which vehicles arrive at a given stop
- Getting the list of stops on route for a specified line number
- Checking the time of arrival in advance based on the timetable
- Route planning by typing the origin stop code and the destination stop code, as a result a list of lines is presented.

All the messages are given by voice, so there is no need to have text-to-speech software installed on the mobile phone.

The behavior of the system in the case of passengers notifying of arriving vehicles is shown in Fig. 2.

To summarize, the existing solutions partially remove the barriers mentioned before. Usually they require proprietary equipment, which is not in mass production. This often results in a high price of deployment of such systems in cities. We considered it as the main disadvantage. Beside this, the other problem that we wanted to resolve is the provision of the information about arriving vehicles and the location of a stop in a way that is acceptable to the visually impaired and blind people. The proposed system is described in the next chapter.

5 Evaluation of the System Proposed

During the tests in cities it has been proved that the proposed system is working properly providing up-to-date information about arrivals of public transport vehicles and giving access to the schedule. However, it became obvious that some technical solutions should be altered to eliminate issues inconvenient for the users.

The main problem the users faced when using the system was the necessity of connecting to the internet in order to get the most up-to-date information about bus arrivals. Another problem was finding the location of the bus stop and getting its unique identifier. The proposed solutions: i.e. using a GPS receiver (built-in or external) or finding a sticker with a stop code written in Braille were not satisfactory because of the following reasons: the first is that only a very small number of mobile phones used by the blind people have had built-in GPS receivers. Those users who have used external GPS modules find them very unhandy because there is a need of turning them on, keeping battery loaded and so on. Finding a sticker with a code number was not an easy task even in the case when they knew where it should be. Having in mind those inconveniences we decided to introduce some changes to the deployed system.

The main change was to install Bluetooth enabled devices at some bus stops which acted as mediators between the system center and the mobile devices used by the passengers.

In that case those devices are constantly connected to the system center using a GPRS connection and become notified of all events which are relevant for the passengers waiting on that stop. The end users do not have to use a GPRS connection in order to get the information from the system center – it is sufficient to get it from those new devices via a Bluetooth interface.

Since all Bluetooth devices have unique physical addresses, the detection of one of them uniquely determines the bus stop where the user is and in consequence eliminates the necessity of getting this information using the GPS or by finding a sticker with a bus stop written in Braille.

6 Conclusion

The proposed system solves the most important problems faced by the blind and visually impaired passengers when they are using public transport means. It provides access to the most up to date information such as the number of the bus which is arriving at the stop. Moreover it is easy to use, it does not require any proprietary equipment, but ordinary mobile phones. All the messages are given by voice. It has been successfully tested in two Polish cities: in Warsaw, at the City Hall subway station and in the public transport company in Nowy Sącz. After the deployment, some drawbacks, like the cost of the data transmission which have to be covered by the end users, appeared to be a problem, even that some GSM operators provided special rates for blind people and the amount of data exchanged between mobile devices and central server was very small.

After additional research we decided to use a short-range wireless communication medium which gave an access to the system center in a range of the bus stop without the necessity of being connected to the system center via a GPRS connection by the end users and eliminated the problem of determining the location of the user. This enhancement improved the overall usability of the system.

References

1. Hersh, M.A., Johnson, M.A. (eds.): Assistive Technology for Visually Impaired and Blind People. Springer, Heidelberg (2008)
2. Miesenberger, K., et al. (eds.): ICCHP 2006. LNCS, vol. 4061. Springer, Heidelberg (2006)
3. Sanchez, J., Maureira, E.: Subway mobility assistance tools for blind users. In: Stephanidis, C., Pieper, M. (eds.) ERCIM Ws UI4ALL 2006. LNCS, vol. 4397, pp. 386–404. Springer, Heidelberg (2007)
4. Załącznik nr XXXV/463/08 Rady Miasta Krakowa z dnia 27 lutego 2008 r. do uchwały w sprawie: przyjęcia programu dostosowania komunikacji miejskiej do obsługi osób niepełnosprawnych (2008)
5. Pięta, N., Szczygieł, M., Wójcik, K., Skąpska, K.: Projekt "Niewidzialne Miasto"– Raport z badań, Fundacja Centrum Analiz Regionalnych, Kraków (2009)
6. Landau, S., Bourquin, E., Miele, J., Van Schaack, A.J.: Demonstration of a universally accessible audio-haptic transit map built on a digital pen-based platform. In: Proceedings of the Third International Workshop on Haptic and Audio Interaction Design, Jyväskylä, Finland (2008)
7. Wiethoff, M., et al.: Specification of Information Needs for the Development of a Mobile Communication Platform to Support Mobility of People with Functional Limitations. In: Stephanidis, C. (ed.) UAHCI 2007 (Part II). LNCS, vol. 4555, pp. 595–604. Springer, Heidelberg (2007)

Pilot Testing of Certification Method Developed for ITS Applications Using GNSS Systems

Miroslav Svitek[1], Vladimir Faltus[1], and Zdenek Lokaj[2]

[1] CTU Prague, Faculty of transportation Sciences,
Konviktska 20, 11000 Prague, Czech Republic
{svitek,faltus}@fd.cvut.cz

[2] e-Ident – Laboratory for Electronic Edentification systems and Comms,
Na Zertvach 34, 18000, Prague Czech Republic
lokaj@e-ident.cz

Abstract. Project of applied research called "Analysis of technical and metrological requirements for GNSS receiving devices and their operations" is a part of the Metrology development program funded by the Czech Office for Standards, Metrology and Testing, started in 2008. The second part of the project executed in 2009 was focused on certification procedures preparation for first ITS applications and pilot testing of the procedures. The virtual gate passage detection (used in the systems such as the EFC) was chosen as one of the tested applications. Results will be used as a benchmark for later testing and certification. The goal of the project consists of the certification authority establishment to provide verification of defined parameters achievement in ITS critical systems which increases transport effectiveness and quality.

Keywords: GNSS, ITS, certification, simulation, on board unit.

1 Introduction

Preparation of certification methodology is in general based on required system parameters (e. g. accuracy, reliability) definition and the process of verification that the application meets these requirements. The definition of system parameters' values, which will be required for successful certification, is one of the most difficult parts of this project. For that reason there have to be executed many tests and pilot subprojects to get sufficient set of data and define these values.

Pilot testing has been executed during Q3 and Q4 2009 in the e-Ident laboratory in Prague. Two ITS applications were tested, the virtual gate passage detection and the passed road segment detection, this article discusses the first tested application. For GNSS simulation the Spirent GSS 8000 simulator has been used and as a reference GNSS system the GPS system was chosen. This way of testing allows defining initial conditions and managing the influence of conditions during the test execution.

1.1 Certification System in General

Certification methodology for each ITS application is based on a general methodology for application evaluation and appropriate system parameters guarantee. The list

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 278–285, 2010.

of representative telematic performance indicators was developed and widely accepted in the structure 0: accuracy, reliability, availability, continuity, integrity and safety. These parameters affect overall ITS system efficiency which should be assessed based on evaluation methodologies as described in 0 and 0.

A general methodology for application evaluation and telematic system parameters guarantee approach is figured bellow on the scheme.

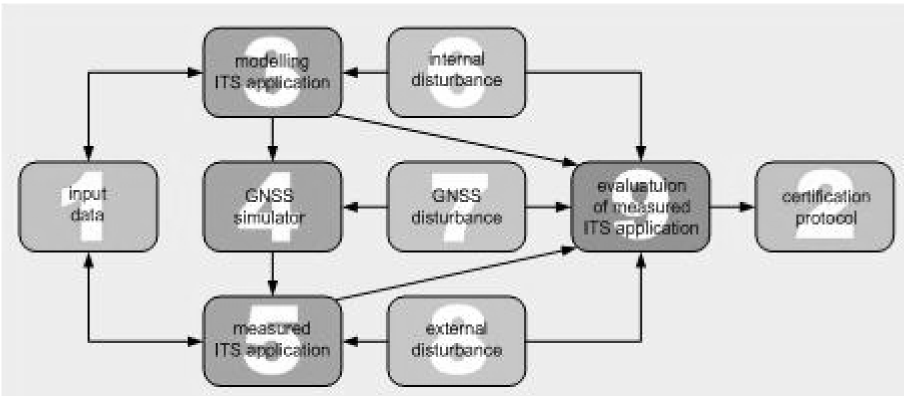


Fig. 1. Certification system for ITS applications evaluation with the use of GNSS simulator, in general

ITS application certification process consists of the following steps:

1) Definition of initial conditions for ITS application (input data – block 1 in Fig. 1);
2) Definition of optimally operating (modelling) ITS application (block 3);
3) Definition of disturbance statistics – internal disturbance of ITS application (block 6, disturbance of the whole set of tested vehicles or OBUs), disturbance of GNSS signal (block 7) and external disturbance (block 8);
4) Activation of measured (certified) ITS application (block 5);
5) Testing of measured ITS application with simulated GNSS signal (block 4) for all defined situations/scenarios (initial conditions – block 1) with the goal to cover all suitable situations;
6) Real testing of selected (available) scenarios;
7) Conformity assessment of output data from appropriate (tested) ITS application and output data of model (optimally operating) ITS application for defined initial conditions and defined disturbances;
8) Measurement results processing for the certification protocol (protocol of measurement) and final assessment of ITS system parameters guarantee (block 2); ITS system parameters guarantee must be statistically verified on a sufficient number of measurements to be able to guarantee monitored properties in defined statistical parameters.

1.2 Certification System in General

Identification of the virtual gate passage was selected as one of the ITS applications tested in the pilot tests. The goal of this certification is to provide safe and guaranteed position and time identification of the vehicle (or human or goods) with an on board unit (OBU).

Virtual gate (portal) is a specific cross profile of the road, where we identify the vehicle passage by means of information from GNSS systems e. g. for fee collection. This term is derived from a physical portal (gate), e. g. equipped with devices for DSRC communication for vehicle identification within EFC systems operated in the Czech Republic, Austria etc. Thanks to virtual gates usage it is possible to expand ground (terrestrial) toll systems by a corresponding system which does not require portals construction.

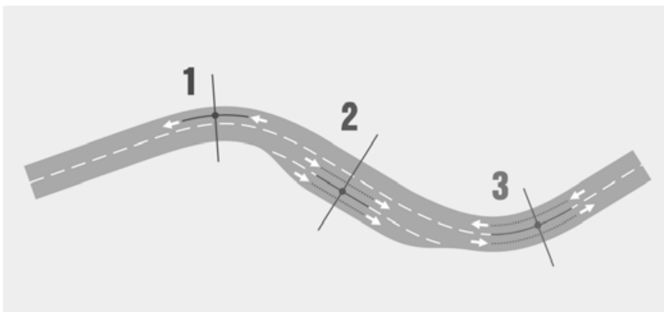


Fig. 2. Definition of virtual gates passage points system

The reference point of the virtual gate passage can be defined as a cross point of traffic lane axis and virtual gate plane. Fig. 2 shows the position of passage points, we can think over one-way (case 1 – traffic lane or case 2 – group of traffic lanes) or two-way (case 3 – one point in the road axis is defined for two directions). If we specify passage points directionally (case 1 and 2), we ensure higher fixing accuracy regardless of GNSS systems accuracy.

Virtual gates are defined by the location (geographical coordinates) and direction (azimuth of drive), which means that the defined virtual gates are one-way. In case of a two-way gate need, two gates with the same coordinates are defined, but with different azimuth.

2 Pilot Testing in 2009

Preparation of certification. The pilot tests of GNSS signal reception conditions and their evaluation were carried out during the months of September to November 2009 in the e-Ident laboratory in Prague. The signal was simulated with Spirent GNSS Simulator, GSS 8000 type. For pilot testing the GPS positioning system was selected as a reference GNSS system.

The routes were generated by creating NMEA messages (a special file was created for each route), every case in the area of Prague and its nearby neighbourhood. Logs of routes for pilot testing were created during September in real test rides in cars, with GPS units. This data was then processed and adjusted according to requirements for pilot testing.

For the virtual gate passage tests, the hardware equipment of universal telematic mobile unit was used as an OBU, developed by Honeywell, Telematix Software, Telematix Services and CTU Prague, Faculty of Transportation Sciences, within the Ministry of Industry and Trade project. On a testing hardware device, the DEFT applications (Dynavix, EFC, Fleet, Toll) have been installed, also developed within the above project, which have been specially adapted for testing needs.

2.1 The Virtual Gate Passage Detection Testing

For the testing, two road sections west of Prague were chosen. The virtual gate passage detection testing (with usage of GNSS simulator) was carried out during the months of October and November 2009 at a total of 26 series of 10 measurements:

- 15 series for the section of road II/605, i. e. a total of 150 measurements,
- 11 series for the section of highway R5, i. e. a total of 110 measurements.

The tests were carried out at first with the default GNSS signal without any restrictions and with the defined parameters of the route. After that further test scenarios were developed and the simulated signal was influenced in the following way:

- changing the signal power from different satellites;
- turning off selected satellites;
- simulating built scenes of surroundings – highway, city, suburbs – which contain a typical set of ground clutter, multipath signal transmission and signal shading;
- simulating various predefined atmospheric changes.

The signal influence has been prepared based on a detailed description of possible signal influence in 0. For the tests a spherical characteristics of the receiving antenna was chosen, with an open top of the antenna (i. e. position of the antenna on the roof of the vehicle).

2.2 Results of the Virtual Gate Passage Detection Testing

The tests performance has been recorded in the “log file” stored directly in the hardware unit. Based on this data further processing took place already on PC. The hardware unit recorded the following data of a virtual gate passage:

- virtual gate passage time;
- ID and a description of the gate (defined positions and azimuth).

Each measurement result was then classified as “passed” or “failed” according to the following categorization:

- passed, if all the following conditions are true:
  - gate is identified on the running road;
  - gate is not identified in other nearby road;
  - any upstream gate is not identified;
  - more passages through the same gate during one test are not evaluated;
- failed in other cases.

Results of tests carried out are listed in Table 1 below.

Table 1. Results of pilot testing, depending on the segment, direction and type of test

Section and direction	Type of test	Passed in %	Failed in %
Road II/605 forward	Passed gate	60.0	40.0
	Upstream gate	100.0	0.0
	Nearby downstream gate	97.3	2.7
	Nearby upstream gate	98.7	1.3
Road II/605 backwards	Passed gate	50.7	49.3
	Upstream gate	98.7	1.3
	Nearby downstream gate	96.0	4.0
	Nearby upstream gate	100.0	0.0
Highway R5 forward	Passed gate	65.5	34.5
	Upstream gate	96.4	3.6
	Nearby downstream gate	100.0	0.0
	Nearby upstream gate	100.0	0.0
Highway R5 backwards	Passed gate	69.1	30.9
	Upstream gate	100.0	0.0
	Nearby downstream gate	100.0	0.0
	Nearby upstream gate	100.0	0.0
On the whole	Passed gate	60.4	39.6
	Upstream gate	98.8	1.2
	Nearby downstream gate	98.1	1.9
	Nearby upstream gate	99.6	0.4

Based on the measurements, it was demonstrated, that the results of the passage identification varies for different parameters of the GNSS signal, environment and other influences on the signal reception. The relatively low percentage of successful running gate identification may be due to high sensitivity software in the OBU. It can be assumed that to use OBUs for ITS applications a higher success rate for negative detection of upstream or neighbouring gates will be demanded.

2.3 System Parameters of Tested Applications

Within the implementation of a sufficiently large count of measurements the resulting values in Table 1 show the probability of conformity of the tested OBU properties with the desired properties of the measured ITS applications.

The proposed test protocol is divided into a part of recording the individual measurements and a part of evaluation of the application as a whole, including the evaluation of system parameters. On the basis of pilot testing it is already possible to summarize the partial results and define partial requirements for selected system parameters of tested applications:

a) Accuracy – the required value of 15 meters in the horizontal plane at 95 % level of probability. Thus the defined accuracy corresponds to usual accuracy requirements for OBU GPS units using a standard statistical distribution of signal parameters. The tested OBU and its software configuration were selected based on the required accuracy. Accuracy of the OBU can be further increased.
b) Reliability – the specific desired value will be defined on the basis of experience in the follow-up testing. We are assuming a value of around 90 %, eventually higher, at 95 % probability level. After pilot tests' result with 60.4 % success of virtual gate passage detection, it can be seen that a higher reliability will be required for the approval. To approve the usage of tested OBU in the EFC system it will be necessary to adjust the software in the OBU or to use another OBU. The reliability is affected by the security parameter, see below.
c) Availability – the specific desired value will again be defined on the basis of experience in further testing. We are assuming the OBU unit activation at the start-up of travel and the availability value up to 60 seconds. Pilot testing showed in most cases the availability of tens of seconds.
d) Continuity – for the EFC this is not a critical system parameter. It depends on the ITS system ability to assess and calculate the travelled route in case of system failure while driving the vehicle; this parameter does not depend only on the OBU and GNSS signal reception. In the case of more virtual gates usage on the road sections the requirements for this parameter are significantly decreasing.
e) Integrity – this parameter has not been considered for the tested systems, mainly because it represents rather the quality of OBU diagnostics, informing the user within a reasonable time for failure of proper OBU function. This parameter is not so much related to position determining or frequency of the virtual gates passage records.
f) Safety – "dangerous conditions" are defined for both tested applications in the field of EFC. This is a situation where the ITS system assesses charges, which in fact did not occur. This is part of the conditions to test status "failed" – so-called "false alarms" – i.e. identification of neighbouring gates or division, identification or evaluation of multiple transits of the same gate. With regard to safety (as one of the important system parameters of tested ITS applications) appropriate testing scenarios were selected by pilot tests – all the scenes are containing unpleasant situation in a similar gate on nearby parallel road (i.e. identically oriented and situated at a distance of tens of meters from the passing road). It should be emphasized that the effort to eliminate false alarms is associated with a reduction of the reliability parameter of ITS applications, i.e. reducing the probability of passed gates correct detection. It is therefore necessary to seek to balance, which on the one hand significantly reduces the risk of false positives, on the other hand, provides a useful reliability parameter. This balanced condition can be found by testing various OBU units, by testing one OBU at various software settings.

Final desired values of system parameters will be determined after the completion of further testing at the end of 2010.

2.4 Conclusions Leading to a General Procedure for the Assessment of ITS Systems Using GNSS Systems

According to the results of pilot tests, the claims on performing the other tests can be determined. Based on results from pilot testing last year and further testing this year it will be possible to summarize the overall certification process.

In other tests a crucial role will be played by the choice of an appropriate mix and number of tests for individual signal characteristics and surrounding, with appropriate repetition of random errors with a given probability distribution at the input to the tested applications.

It is necessary to retreat from built-in scenes of surrounding in the simulator and to define own scenes based on research and development 0. It is also necessary to test other external and internal influences on GNSS signals, as well as to test multiple gates for one test – concordant, counter or in other ways (e.g. by perpendicular or oblique crossing of roads). Results of testing should then be interpreted depending on characteristics of GNSS signal and influence on GNSS signal.

The aim is to find the optimal test scenario from the perspective of testing effectiveness, especially the model of disturbances. Selected testing results can be verified by the field measurement.

In case of unsatisfactory results, it can be considered in the future how to improve the accuracy of determining the position by using other systems:

- usage of other GNSS system in addition to the GPS system – such as GLONASS or European Galileo system;
- usage of SBAS system – e. g. the European EGNOS system;
- usage of GBAS system – e. g. CZEPOS system operated in the Czech Republic by the Czech Office for Surveying, Mapping and Cadastre; this system provides users with a GPS correction data for accurate positioning in the Czech Republic; CZEPOS is managed and operated as part of the geodetic foundations of the Czech Republic.
- Usage of non-direct positioning systems, e.g. inertial system, odometer, vehicle sensors etc.

3 Conclusion

The previous analysis and findings from the pilot testing in 2009 are the basis for the research this year – further testing of vehicle units and ITS applications under various operating conditions. The tests and certification methodology will be supplemented by other ITS applications with higher demands on the system parameters, such as critical applications for public transport according to the project called “Telematic support for sustainable transport development in the regions”. Attention is paid to individual components of ITS applications, as well as to entire ITS subsystems. A precise methodology for certification of specific ITS applications will be based on the results of the new testing.

Certification authority establishment as a result of ongoing research will allow the use of GNSS systems in the critical ITS applications which in fact will increase transport safety, effectiveness and quality. System parameters values and testing methods definition are crucial in this process.

References

1. Svitek, M.: Dynamical Systems with Reduced Dimensionality. In: Neural Network World edn., II ICR AS CR and Faculty of Transportation Sciences, CTU Prague, Prague (2006) ISBN: 80-903298-6-1, EAN: 978-80-903298-6-7
2. Svitek, M., Faltus, V., Lokaj, Z.: Analysis of Technical and Metrological Requirements for GNSS Receiving Devices and Their Operations, Annual report, Prague (2009)
3. Puricer, P., Kacmarik, P., Kovar, P., Spacek, J., Vejrazka, F.: Analysis of GNSS signal received by mobile user. In: Faculty of Electrical Engineering, CTU Prague, Research study, Prague, pp. 26–32 (2009)
4. Svitek, M.: Intelligent Transport Systems – Architecture, Design methodology and Practical Implementation. In: Key-note lesson, 5th WSEAS/IASME Int. Conf. on Systems Theory and Scientific Computation, Malta (2005)
5. Mikulski, J. (ed.): Advances in Transport Systems Telematics. WKŁ, Warszawa (2008) ISBN: 978-83-206-1715-3; Stárek T., Svítek M.: Practical examples of ITS evaluation, ch. 36, pp. 295 – 306
6. Mikulski, J. (ed.): Advances in Transport Systems Telematics. Monograph. Silesian University of Technology, Katowice (2006) ISBN: 83-917156-4-7; Svítek M., Stárek T.: ITS evaluation methods, ch. 4, Section I – Telematic transport systems – general issues, pp. 33–43

Measuring Transducer Modelled by Means of Fractional Calculus

Mirosław Luft, Elżbieta Szychta,
Radosław Cioć, and Daniel Pietruszczak

Kazimierz Pułaski Technical University of Radom,
Faculty of Transport and Electrical Engineering,
26-600 Radom, Malczewskiego 29, Poland
{m.luft,e.szychta,r.cioc,d.pietruszczak}@pr.radom.pl

Abstract. The article is inspired by developments of the fractional calculus in different areas of science such as the control theory and electrical measurements. The current interest in mathematical analysis employing the fractional differential and integral calculus reflects the usefulness of this calculus in the development of more precise – closer to the actual observation - mathematical models of various phenomena. A model of a measuring transducer is presented, developed by means of fractional calculus. Tests are executed in the programming environment MATLAB-SIMULINK.

Keywords: fractional calculus, measurement transducer, dynamic systems, discrete transmittance.

1 Introduction

Derivative integral calculus is nothing new. It originated in the 17th century. It was mentioned in G. W. Leibnitz's letters to de l'Hospital (1695), in works by L. Euler (1738) or P. S. Laplace (1812) [8], [14], and [16].

A range of physical effects, such as the penetration of a liquid through porous substances, penetration of charges through real insulators or penetration of heat through a heat barrier are more accurately described by means of derivative integral equations. Dynamics of such physical processes as acceleration, displacement, liquid flow, intensity of electrical current or magnetic flux are modelled employing differential equations. These processes are in fact $m+1$ times differentiable continuous variables, where m depends on the order of a non-integral variable under examination. It is impossible to shift a mass from one location to another over an infinitely short time or to change the temperature or pressure of a real object at an infinitely fast rate [16].

The dynamic growth of research into applications of fractional order derivative integral calculus to the analysis of dynamic systems in recent years motivates attempts at its employment in modelling of measuring transducers [11] and [16].

The classic notation of a measuring transducer dynamics is based on differential equations, which are a mathematical model of the transducers in the domain of time [1], [2], [3], [4], [6], [7], and [10].

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 286–295, 2010.

An equation of this type is given:

$$\begin{aligned} &A_i \frac{d^{(i)}y}{dt^{(i)}} + A_{m-1}\frac{d^{(i-1)}y}{dt^{(i-1)}} + ... + A_0 y(t) = \\ &= B_j \frac{d^{(j)}f(x)}{dt^{(j-1)}} + B_{m-1}\frac{d^{(j-1)}f(x)}{dt^{(j-1)}} + ... + B_0 f(x) \end{aligned} \tag{1}$$

where: $y = f(x)$.

In derivative integral calculus, function differentiation and integration operators are combined into a single operator D^n. In respect of differentiation, the order n becomes positive $n = 1, 2, 3, ...$, while with regard to integration, it is negative $-n = -1, -2, -3, ...$. A neutral operator for $n = 0$ is also defined:

$$D^n f(t) = \begin{cases} \dfrac{d^n f(t)}{dt^n} & \text{for} \quad n > 0 \\ f(t) & \text{for} \quad n = 0 \\ \int\limits_{t_0}^{t}\left[\int\limits_{t_0}^{\tau_1}...\left[\int\limits_{t_0}^{\tau_{-n-1}} f(\tau_{-n})d\tau_{-n}\right]...d\tau_2\right]d\tau_1 & \text{for} \quad n < 0 \end{cases} \tag{2}$$

The classic engineering approach assumes only that $t_0 \geq 0$.

In fractional order derivative integral calculus, a derivative of any order is treated as an interpolation of a sequence of discrete order operators (2) by means of continuous order operators. The notation introduced by H. D. Davis is employed [5], where a derivative of a non-integral function $f(t)$ can be expressed:

$$_{t_0}D_t^{\nu} f(t) \tag{3}$$

where:

t_0, t - terminals of fractional differentiation or integration;

ν - order of a derivative integer.

It should be noted that (3) defines the same differentiation range $[t_0, t]$, as in classic definite integrals. Therefore, a non-integral order derivative and integral are defined in the range $[t_0, t]$, which, in the case of the derivative, narrows down to a point (range) $[t, t]$ for the integral order n. To distinguish between integral and non-integral orders, the latter are designated with Greek letters ν, μ. The commonly used letters m and n are reserved for integral orders.

An order of a derivative or an integral fulfils the condition:

$$\nu \in \mathbf{G}_+, n \in \mathbf{Z}_+ \tag{4}$$

where:

$\mathbf{G}_+$ - set of positive real integers; $\mathbf{Z}_+$ - set of non-negative integers.

To highlight the difference between a derivative and an integral, H.D. Davis's notation [5] comprises the expression $_{t_0}D^{-\nu} f(t)$, where the integration order meets the condition (4), and the minus by D indicates this is an integration.

2 Classic Notation of the Model of Seismic Mass Measuring Transducer

In the classic notation, the dynamics of a measuring transducer design shown in Fig. 1 is expressed as a 2nd order differential equation [1], [6], and [11]:

$$\ddot{w}(t) + 2\zeta\omega_0 \dot{w}(t) + \omega_0{}^2 w(t) = -\ddot{x}(t) \tag{5}$$

where: ω_0 – natural angular frequency, ζ – damping.

Seismic mass transducers, depending on their execution (selection of parameters characterising their dynamic properties), can be used to measure displacement, speed or acceleration. The displacement $x(t)$ is the input quantity in these transducers. In practical vibration measurements, transducers are applied to acceleration measurements. The parameters of speed and displacement are determined by applying elements that integrate accelerometer signals.

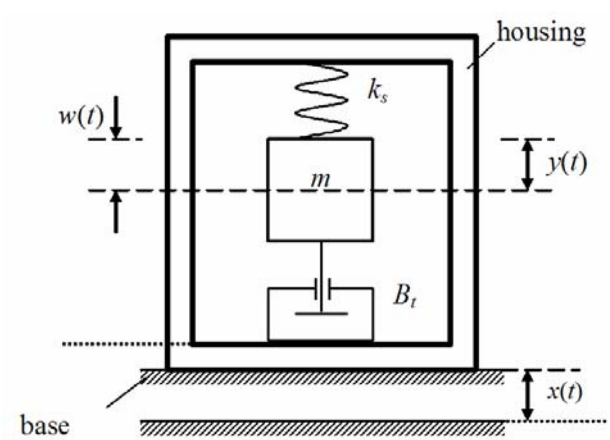


Fig. 1. Model of a seismic mass measuring transducer [1]

The transducer comprises: m – transducer seismic mass, k_s – reaction of spring, B – transducer damping, $x(t)$ – displacement of base and housing, $y(t)$ – displacement of mass m in relation to the fixed coordinate system and $w(t)$ – displacement of mass m towards the base.

Equation (5) can be noted as a difference equation:

$$\begin{aligned} &a_2 w_k + a_1 w_{k-1} + a_0 w_{k-2} = \\ &= b_2 x_k + b_1 x_{k-1} + a_0 x_{k-2} \end{aligned} \tag{6}$$

or a matrix equation:

$$\begin{aligned} &\begin{bmatrix} a_2 & a_1 & a_0 \end{bmatrix} \begin{bmatrix} w_k \\ w_{k-1} \\ w_{k-2} \end{bmatrix} = \\ &= \begin{bmatrix} b_2 & b_1 & b_0 \end{bmatrix} \begin{bmatrix} x_k \\ x_{k-1} \\ x_{k-2} \end{bmatrix} \end{aligned} \tag{7}$$

The derivative-integral expression of (6) becomes:

$$\begin{aligned} &A_2 {\Delta_k}^{(2)} w_k + A_1 {\Delta_{k-1}}^{(1)} + A_0 w_{k-2} = \\ &= B_2 {\Delta_k}^{(2)} w_k + B_1 {\Delta_k}^{(1)} x_{k-1} + B_0 w_{k-2} \end{aligned} \tag{8}$$

where $\Delta_k^{(n)}$ is the reverse difference of the discrete function, defined as:

$$\Delta_k^{(n)} f_{(k)} = \sum_{j=0}^{k} a_j^{(n)} f_{(k-j)} \tag{9}$$

Once (9) is taken into consideration, (8) noted as a matrix equation becomes (10):

$$\begin{aligned} &\begin{bmatrix} a_2 & -a_1 - 2a_0 & a_2 + a_1 + a_0 \end{bmatrix} \begin{bmatrix} \Delta_k^{(2)} w_k \\ \Delta_k^{(1)} w_k \\ \Delta_k^{(0)} w_k \end{bmatrix} = \\ &= \begin{bmatrix} b_0 & -b_1 - 2b_0 & b_2 + b_1 + b_0 \end{bmatrix} \begin{bmatrix} \Delta_k^{(2)} x_k \\ \Delta_k^{(1)} x_k \\ \Delta_k^{(0)} x_k \end{bmatrix} \end{aligned} \tag{10}$$

3 Analysis of Response of the Derivative-Integral Model of a Measuring Transducer

Responses of a measuring transducer to an input sinusoidal signal are compared and described with the aid of three models [11]:

1. Continuous-time model noted by means of an operator transmittance:

$$G(s) = \frac{-s^2}{s^2 + 51s + 255} \tag{11}$$

2. Discrete model, derived from the continuous-time model noted by means of a discrete transmittance:

$$G(z) = \frac{-z^2 + 2z - 1}{z^2 - 1.975z + 0.9748} \tag{12}$$

3. Discrete model, determined by a derivative-integral notation obtained from (10) and (12):

$$G(z) = \frac{-z^2 + 0.02524z - (6.294e-005)}{z^2 - (3.161e-005)z + (1.11e-016)} \tag{13}$$

Responses of the models were tested in the programming environment MATLAB [17]. An irreplaceable work tool here was SIMULINK, an interactive package built on the basis of MATLAB. It offers a possibility of analysis and synthesis of continuous and discrete dynamic systems. SIMULINK is a graphic environment where the dynamic system simulation is accomplished on the basis of a block diagram built with the use of library blocks. Fig. 2 presents a flow diagram of the measuring system.

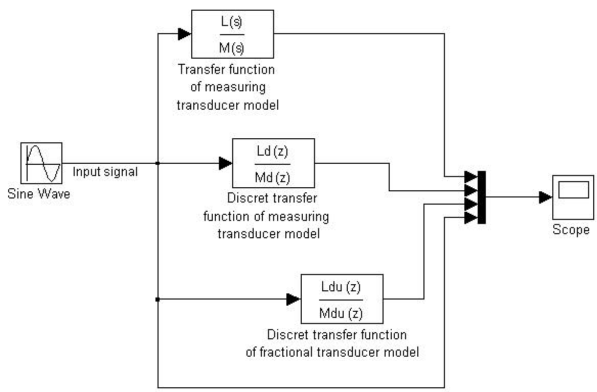


Fig. 2. Flow diagram of the measuring system

Model (12) was obtained by discretising the continuous model (11) using the 'zero-order hold' method (Fig 3) at a sampling time of $T_p = 0.0005s$. Zero-order hold (ZOH) devices convert sampled signals to continuous-time signals for analysing sampled continuous-time systems. The ZOH discretisation $H_d(z)$ of a continuous-time LTI model $H(s)$ is depicted in the following block diagram:

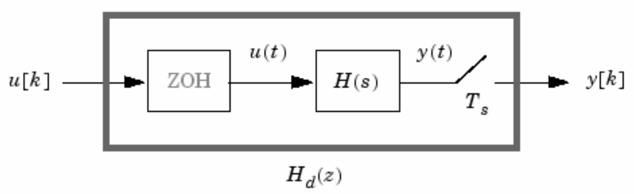


Fig. 3. Zero-order hold method [17]

Responses of all the models to a 100 rad/s input sinusoidal signal are presented in Fig. 4. Signals from models 1, 2 and 3 are phase-shifted in relation to the input signal. It should be pointed out that model 3, described by means of transmittance (13), correctly reproduces the value of the input signal amplitude.

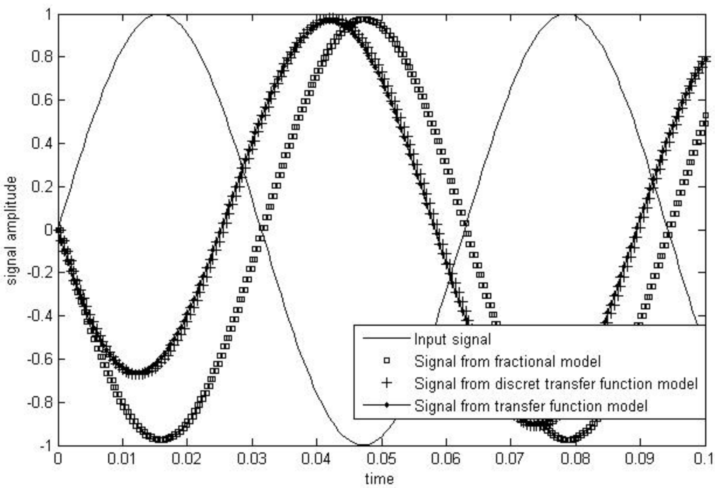


Fig. 4. Comparison of responses of measuring transducer models

Fig. 5 shows frequency Bode diagrams for model 1, described by means of transmittance (11). It proves that a stable amplitude gain is accomplished starting with approximately 100 Hz, with a concomitant phase shift from 180° to 200°.

Fig. 6 compares frequency Bode diagrams for the discrete model of measuring transducer (12) and the discrete model of measuring transducer as determined by derivative-integration notation (13).

The comparison of the model's responses to the sinusoidal input signal (Fig. 6) demonstrates that the derivative-integration model of measuring transducer (13) correctly reproduces the amplitude of the input signal from the beginning of the simulation. The classically determined model (12) correctly reproduces the amplitude of the input signal after the end of non-stationary state – after 0.02 s in our testing.

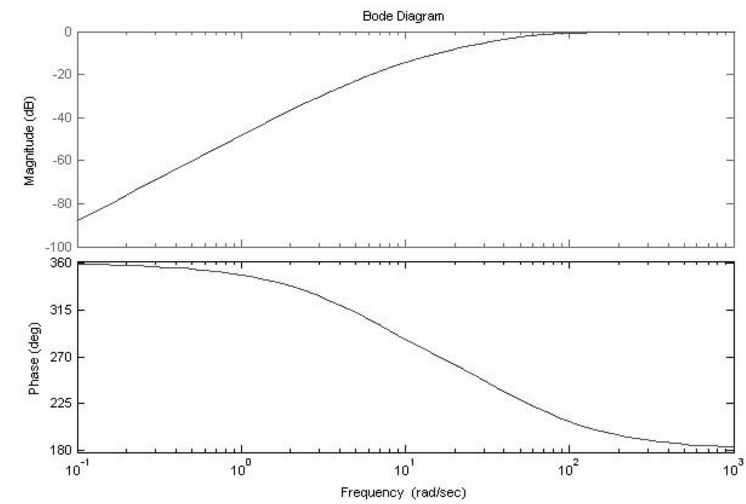


Fig. 5. Bode diagrams for model (11)

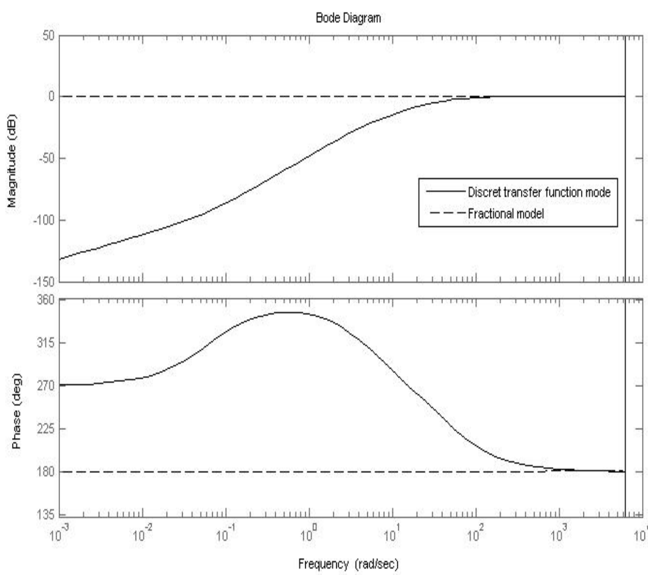


Fig. 6. Comparison of Bode diagrams for measuring transducer models (12) and (13)

It can be concluded from the Bode diagrams (Fig. 6) that the processing scope of the input signal amplitude is extended to include low frequencies in the case of the measuring transducer model obtained by derivative-integration methods as compared to the classically determined model. In the case of the presented diagram, 0 dB gain in the amplitude of the derivative-integration model is obtained starting with frequencies of 0.001 rad/s, compared to 100 Hz in the 'classic' model, with a constant phase displacement of 180°.

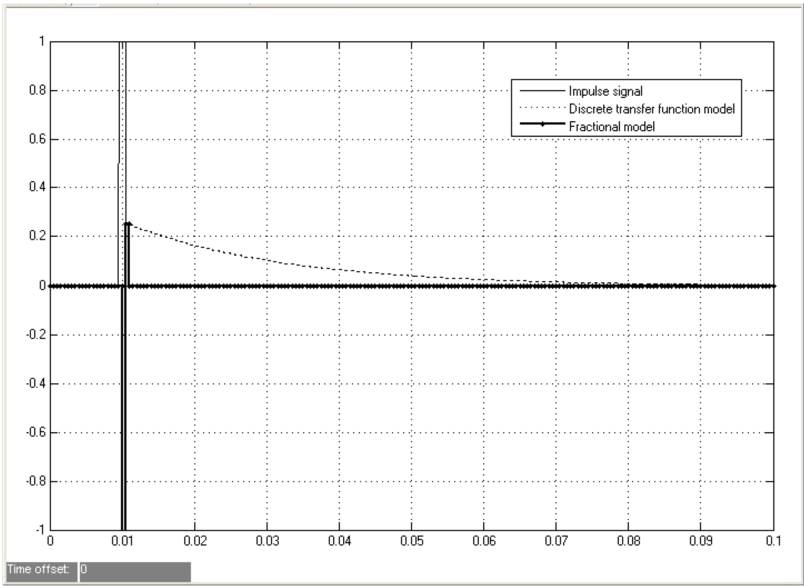


Fig. 7. Comparison of pulse function measuring transducer models

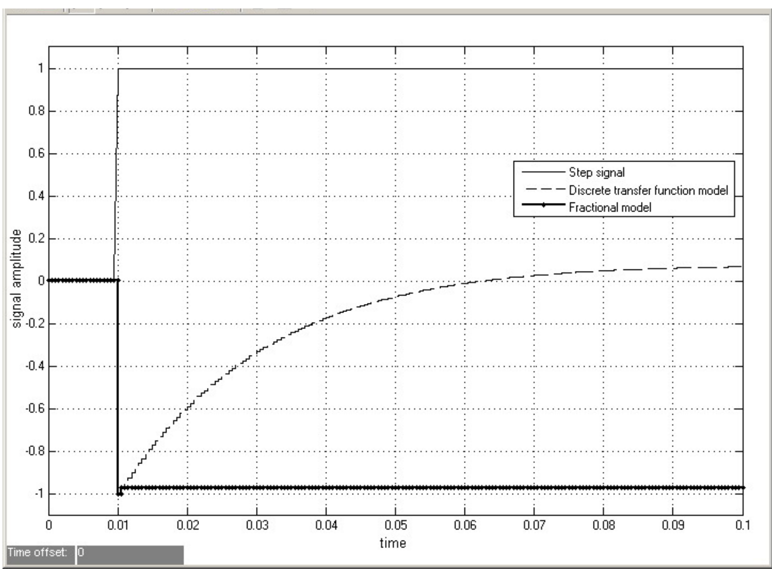


Fig. 8. Comparison of step function measuring transducer models

Fig. 7 compares responses of the models presented to pulse functions. The response of the classic model enters its steady state 0.07 s after the occurrence of the signal. With regard to the derivative-integration model, the same time is reduced to 0.001 s. Fig. 8 compares responses of the models presented to step functions.

The response of the classic model enters its steady state 0.6 s after the occurrence of the signal. With regard to the derivative-integration model, the steady state is obtained after 0.005 s and its value is approximately equal to that of the reverse sign pulse amplitude.

4 Conclusions

Differential and integral calculus involving fractional, or possibly real or integrated order derivatives has become extremely popular. Professor Katsuki Nishimoto said that "Fractional calculus will become the calculus of the 21st century" [12]. It constitutes a response to the 'classic' mathematical apparatus which produces less precise mathematical models. A derivative and an integral of any orders open as yet unimaginable possibilities for the identification of dynamic systems, creation of new, previously unfeasible control algorithms in feedback systems. Orders can also be seen as functions of time. This produces differential equations of variable, time dependent orders.

Applying the derivative-integral method to the development of models of vibration measuring transducers results in an input signal processing model that is ideal in the case of amplitude reproduction [11]. In the case of pulse signals that are invariable over time and of amplitudes other than zero, the response of the derivative-integral model is approximately equal to the negative amplitude of the pulse signal. In conclusion, the modelling of measuring transducers by the derivative-integral method is more accurate than the classic differential equation method when the dynamics of a measuring transducer response is under examination. It should be pointed out that the commonly known derivatives are merely special cases of the calculus presented in this article. An integral order of an integral is understood to correspond to a rate of definite integral. One should therefore speak of the differential and integral calculus of an integral and non-integral, or random, order.

Further research is required to verify how a transducer model determined by means of the derivative-integral method represents an actual measuring transducer and whether it is more accurate at reflecting the dynamics of the input signal processing than the model described by the 'classic' differential equations.

References

1. Cioć, R.: Korekcja charakterystyk dynamicznych przetworników pomiarowych w diagnostyce wibracyjnej wagonu kolejowego, Ph.D. Thesis, Politechnika Radomska, Radom (2007)
2. Cioć, R., Luft, M.: Valuation of software method of increase of accuracy measurement data on example of accelerometer. In: Advances in Transport Systems Telematics, Monograph, Faculty of Transport, Silesian University of Technology, Katowice (2006)
3. Cioć, R., Luft, M.: Correction of transducers dynamic characteristics in vibration research of means of transport – part 1 – simulations and laboratory research. In: 10th International Conference "Computer Systems Aided Science, Industry and Transport", Transcomp 2006, Zakopane, vol. 1 (2006)

4. Cioć, R., Luft, M.: Metoda programowej korekcji dynamicznych błędów przetwarzania przetworników pomiarowych. Pomiary Automatyka Komputery w gospodarce i ochronie środowiska, Kwartalnik Naukowo-Techniczny nr 2/2009, str. 22-25,, Fundacja Nauka dla Przemysłu i Środowiska, Rzeszów (2009) ISSN 1889-6981
5. Davis, H.D.: The theory of linear operators. Principia Press, Bloomington (1936)
6. Hagel, R., Zakrzewski, J.: Miernictwo dynamiczne. WNT, Warsaw (1984)
7. Jakubiec, J., Roj, J.: Pomiarowe przetwarzanie próbkujące. Wydawnictwo Politechniki Śląskiej, Gliwice (2000)
8. Kaczorek, T.: Wybrane zagadnienia teorii układów niecałkowitego rzędu. Oficyna Wydawnicza Politechniki Białostockiej, Białystok (2009)
9. Kaczorek, T., Dzieliński, A., Dąbrowski, W., Łopatka, R.: Podstawy teorii sterowania. WNT, Warszawa (2006)
10. Luft, M., Cioć, R.: Increase of accuracy of measurement signals reading from analog measuring transducers. In: Zeszyty Naukowe Politechniki Śląskiej 2005,Transport z. 59, Gliwice, vol. 1691 (2005)
11. Luft, M., Cioć, R., Pietruszczak, D.: Measurement transducer modeled by means of classical integral-order differential equation and fractional calculus. In: Proceedings of the 8th International Conference ELEKTRO 2010, Zilina (2010) ISBN 978-80-554-0196-6
12. Nishimoto, K.: Fractional calculus. Integration and differentiation of arbitrary order. Decartes Press, Koriyama (1991)
13. Nishimoto, K.: An essence of Nishimoto's fractional Calculus. Decartes Press, Koriyama (1989/1996)
14. Ostalczyk, P.: Zarys rachunku różniczkowo-całkowego ułamkowych rzędów. In: Teoria i zastosowania w automatyce, Wydawnictwo Politechniki Łódzkiej, Łódź (2008) ISBN 978-83-7283-245-0
15. Mikulski, J., Młyńczak, J.: Wykorzystanie systemu monitoringu GPS do oceny parametrów energetycznych lokomotyw spalinowych. Przegląd Elektrotechniczny 9, 268–272 (2009)
16. Podlubny, I.: Fractional Differential Equations. Academic Press, New York (1999)
17. Matlab®&Simulink®7, The MathWorks™ (2008)

Concept of Integrated Information Systems of Rail Transport

Mirosław Siergiejczyk and Stanisław Gago

Faculty of Transport, Warsaw University of Technology
Department of Telecommunications in Transport
75 Koszykowa St., 00-662 Warsaw
{s.gago,msi}@it.pw.edu.pl

Abstract. This paper will present a need to create integrated information systems of the rail transport and their links with other means of public transportation. IT standards will be discussed that are expected to create the integrated information systems of the rail transport. Also the main tasks will be presented of centralized information systems, the concept of their architecture, business processes and their implementation as well as the proposed measures to secure data. A method shall be proposed to implement a system to inform participants of rail transport in Polish conditions.

Keywords: data exchange, IT standards, information system.

1 Introduction

A rapid growth of freight transport contributes to economic development and employment growth, but also causes congestion, accidents, noise, pollution, and an increased reliance on imported fossil fuels and energy loss. In order to solve these problems, it is necessary to optimise the European transport by applying advanced logistic solutions. The logistics can increase the efficiency of individual modes of transport, as well as their combinations. Co-modality, i.e. the efficient use of various modes separately and in combinations, should become a primary element of logistics, which should result in an optimal and sustainable use of resources [9].

In transport, an important issue is the information flow between the participants in the transport chain. For the purpose of disseminating that information, it is necessary to develop contacts (interfaces) to exchange the data (information) and determine the scope and the means by which various actors involved in the transport process shall access information. This also refers (and perhaps above all) to the railway companies as regards the information flow between the railway infrastructure managers and transportation companies and to the information provided to travellers and users of transport services. Advanced telematic technologies may greatly contribute to the increase in rail transport effectiveness by the improvement to the infrastructure, traffic, and rolling stock management, as well as by changing for the better the identification and tracing of shipments across the transport networks operating areas and by providing relevant information to the transport participants [3], [4], and [5].

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 296–307, 2010.

In order to do that, there is still a need to overcome a number of obstacles, hindering the universal and consistent use of telematic technologies in freight transport logistics, which include insufficient standardisation of the exchange of relevant information and the differentiation of competencies of actors operating on the market in the area of application of those technologies. Matters of security and data privacy protection should also be remembered about.

Among other things, intelligent technologies should be introduced for safety reasons in order to avoid delays in the supply chain. Radio frequency identification (RFID) is one of such technologies and it is a growing market but it requires further research and works on frequencies management, interoperability, and standardisation. Common messaging standards (e.g. EDI / EDIFACT) and new communications platforms (e.g. XML) are further elements of the structure. An attempt to solve those problems leads to the consideration of opportunities to create integrated information systems for purposes of rail transport logistics. Freight transport logistics should be a research priority of the Seventh Framework Programme of the European Commission, because modern technological innovations can open up new gateways for development to that industry [9].

2 Analysis of Needs to Implement Integrated Information Systems of Rail Transport

According to the assumptions, the integrated information systems of rail transport should be the systems informing about the implementation of transport services with the use of rail transport. They should, inter alia, allow for obtaining current information about the issued decisions (licenses, certificates, etc.), information on the transport capacities of national and international rail operators, and serve as a tool to monitor the traffic and check the relevancy of the lawmaking. In Poland, those systems should facilitate the adjustment of the rail transport to the European Union legislation and to create tools for the equitable development of carriers in that market. They should provide direct access to information from fixed and mobile terminals and allow the exchange of information electronically between railway companies, offices, and railway clients [5].

Integrated Information Systems of Railway Transport should support rail transport participants in the area of transportation of passengers, baggage, and goods, in compliance with a description of the subsystems (telematic applications) contained in Appendix II of Directive 2001/16/EC of the European Parliament and the Council of Europe of 19 March 2001 on the interoperability of the trans-European conventional railways system [1].

In accordance with the Appendix to the Directive, the subsystems include two elements:

- applications for passenger services; the applications should include at least the systems to inform passengers before and during the journey, booking and payment systems, baggage management, and management of railway connections and links with other modes of public transport
- applications for freight services; the applications should include at least information systems (real-time monitoring of freight and trains), marshalling and

allocation systems, and reservation, payment, and invoicing systems, management of connections with other modes of transport, and preparation of electronic accompanying documents.

Appendix III to the above Directive defines the requirements that individual subsystems must meet. For the Telematic Applications Subsystem they have been defined as follows:

- in the area of technical compatibility, the primary requirements for telematic applications should guarantee a minimum quality of services for passengers and carriers of goods,
- database, software and data transmission protocols must provide maximum data exchange between different applications and operators, excluding confidential commercial data;
- telematic applications should provide the user with easy access to information,
- in the area of reliability and availability – ways of how to use, manage, update and maintain databases, software, and data transmission protocols must ensure the efficiency of those applications and the quality of services,
- in the area of health – cooperation between the applications and users must comply with minimum rules on ergonomics and health protection,
- in the area of safety – an adequate level of integrity and reliability is required
- in the area of collection and transmission of safety-related information.

3 Information Technology Standards in Integrated Information Systems of Railway Transport

3.1 Information Technologies in Transport

Polish Information Technology Development Strategy until 2013 discusses inter alia prospective information technologies in transport and tourism – e-Transport and e-Tourism. As a result of changes in shaping a new balance between the work time and leisure time new eServices shall appear, particularly in the field of tourism and transport. A group of “rich in money, poor in time” should be growing faster for whom present ways of spending leisure time and travelling will be an unreasonable burden. Transport and tourism services are an area in which the implementation of information technologies will contribute to radical changes over the next 10 years. The dissemination of information technologies in those areas is characterised by three technological factors:

1. Widespread deployment of the new generation smart cards technologies with much more powerful computing capacities and a cache volume higher than the current ones. The implementation of customized functions that are now characteristic of large systems with databases shall become feasible. In the event of transport services, the changes will refer to inter alia:

- Travel payments system – one should expected that "the paper form of tickets" will be eliminated to a large extent. The transport system will become a multimodal one (buying a ticket for "the journey" and not for "the means of transport"). There will be more flexible and individualized offer, taking into account the specific needs of different groups of the services users, in particular with disabilities, e.g. physically disabled.

- Tickets Ordering. Ordering tickets, or rather a journey, will be carried out "from the place of current residence" and a computer system, based on personalized data in the passenger card, will take care not only of the best possible choice of means of transport and routes of the journey, but also that in case of delays or changes during the same trip "the rebooking" would happen in a manner that is "invisible to the passenger". Buying "leisure services" will also be carried in a similar way. The system has a chance of universal implementation in a perspective of 2010-2015, because already at present the chip cards are increasingly used in public transport ticketing systems (e.g. in Warsaw).

2. The development of spatial information systems, which will entail the development of monitoring: global, regional or local. In the perspective of 2013, access to the data of such systems will become widespread. In combination with the development of mobile and satellite technologies, it will create a new quality in planning, organizing, and the course of the journey.
3. The fourth-generation mobile technology 4G, which will also enable low-cost implementation of such solutions as a mobile city guide (inside the phone or a communicator of a tourist) presenting a city visitor with online detailed information, customized to a profile, preferences, and needs of the tourist [7].

3.2 XML Standard in Systems to Provide Services Electronically

The XML (Extensible Mark-up Language) language has become the basic standard in the area of providing services, using electronic platform, as the primary means of data exchange. This standard has been successfully operating among the European Union leaders in the provision of services using the electronic platform. Solutions that are used there already have their designing phase long behind them and operate well, thus contributing to the development of the idea of an information society in those countries. The standard is:

- supported by the IT market, which allows for costs reduction and mitigation of risks associated with the creation of public sector information systems,
- scalable, i.e. susceptible to changes in the requirements for a system, changes in the scope and quantity of information stored, as well as the number of transactions processed, and the number of visitors using the proposed solutions,
- open, i.e. all specifications are documented and made available to all parties concerned.

The use of XML has determined the way of how to create the entire architecture of systems to provide electronic services. The XML standard (language) is used to exchange information between the government systems and between:

- governmental agencies and citizens;
- governmental agencies and institutions intermediating in contacts with a citizen;
- governmental agencies and the business sector (from around the world);
- public sector institutions at lower levels;
- governmental agencies and other central government administrations such as the EU.

In addition, the ebXML standard, which is the solution for the business sector is also developing the concept of electronic document exchange (EDI) using the XML language.

3.3 Standards for Geo-Visualisation

The XML standard is a universal formal language used to represent different data in a structured way. The XML is platform-independent, which allows easy exchange of documents between different systems. The XML is a subset of the SGML language. Different varieties of XML can be used as:

- records of geospatial databases (universal, text, environmentally independent, free);
- geospatial metadata records;
- a communication language (data exchange) between the servers of geo-information services and the client applications (via the intranet or Internet);
- a recording format for the cartographic presentations (maps, visualisations), user applications programming language (e.g., transformations between coordinate systems).

The benefits of using the XML are based mostly on its versatility both as a "carrier" of geospatial data (GML – Geography Mark-up Language), cartographic presentations (e.g. SVG – Scalable Vector Graphics) as well as advanced programming language functionalities of geo-information services. At the present stage of development of geo-information services, the following are of the fundamental importance: dissemination of information systems interoperability, preparation of metadata and directory services, and specialised Web Services. Another significant trend in the development of cartography and dissemination of geo-information are the mobile systems using the location-based technology (LBS – Location-Based Services), including positioning systems (e.g. GPS – Global Positioning System) and mobile telephony devices GSM/GPRS.

3.4 GPS and GSM in Rail Facilities Localisation

The railway administrations currently use various applications using digital communications to quickly search the trains and railway premises along with their visualisation on the map and to obtain the information about current parameters of traction vehicles and trains. The functionality of applications used varies and depends on the additional techniques applied to determine the location, e.g. GPS, EGNOS, inertial sensors, electronic maps of rail routes.

Modern telecommunications, information technology solutions, and satellite navigation allow for providing integrated services that involve, inter alia, determining the location of vehicles, status controlling, and automatic monitoring of cargo.

Mobile telecommunications is now an important part of fleet management systems for both road and rail transport. Opportunities to collect the data about the location of mobile objects in railway transport are virtually limited to two systems: GPS, i.e. Global Positioning System, and GSM. Because of the range, accuracy and cost of equipment, the GPS system is the world leader in terms of localisation. A popular method to collect the data about the location is also the possibility of combining GPS and GSM/GSM-R. Currently, the GSM systems alone allow for identifying the location of trains with great accuracy.

GSM-R networks also enable the provision of localisation services. These services provide the information related to the current location of a subscriber terminal. The biggest advantage of localisation services in GSM-R networks is that the localisation is carried out automatically without user intervention. Such services require the existence of a centre – the server connecting the mobile network to user applications. The localisation requests would be directed to the centre that would measure a geographical location of a mobile terminal and would send feedback out to the user about the position of the mobile terminal. The centre should also ensure storing the data in archives. User applications should be basically equipped with the railway lines electronic maps software.

3.5 Information Systems for Rail Market Participants

Portals

Portal is a place of contact and a principal place of access to electronic public services. It implements many advanced functions such as adapting the interface for an individual user (customization), multilingual character, support for electronic signature and access with the use of alternative devices such as mobile phones.

***CMS* – Content Management System**

CMS (Content Management System) is a software program that allows designing, easy update and efficiently expanding a structure of the existing website. Its support does not require a special IT knowledge. Modifying and adding new materials to the site is practically possible from almost any computer using any web browser WWW.

RSS News Feeds

RSS (Really Simple Syndication, Rich Site Summary) is a web standard for providing information on the changing content associated with the XML (extensible mark-up language). It transmits the news headlines. The full text of notes or their "summaries" are available immediately upon publication; they can be downloaded using a special feed reader. RSS news feeds are also used in transmitting the information between web services.

Feed Reader (news aggregator)

Feed Reader enables entering the news headlines posted in the news channels available on the Internet. The headline contains a title, brief description, and a link referring to more detailed information, contained in a service maintained by the feed publisher. The RSS reader allows for tracking multiple feeds simultaneously.

The advantage of the feed reader is the ability to track online news originating from multiple sources simultaneously in a single interface-date and load those that are interesting for the final reader.

Interactive Maps

Interactive Maps allow for integration with the directory, the database of objects being placed on the map. Features of an interactive map:

- map scalability and its smooth scrolling and retreat/zooming an opportunity to place an unlimited number of points on the map browser search: streets (in the case of a city plan) / village (in the case of maps of the area: the municipality, district, province) / lakes, investment sites, etc.
- a list of objects (stations, bus stops, institutions, firms, historical sites) linked to the map with a possibility to link objects, displayed on the map, with an appropriate description illustrated by photographs and graphics appearing in the directory
- a possibility to attribute individual icons to objects and groups of objects on the map.

4 Concept of Centralised Information Systems in Polish Conditions

The subsystem of Telematic Applications for Freight Transport, described by the Technical Specifications of Interoperability (TSI), is defined in accordance with the requirements of Appendix II of Directive 2001/16/EC. It includes in particular [1]:

- applications for freight transport services, including IT systems (real-time monitoring of cargo and trains),
- marshalling and allocation systems, however, the allocation systems are understood as a train composition,
- booking systems, understood hereto as booking of train routes,
- management of connections with other modes of transport and creation of electronic accompanying documents.

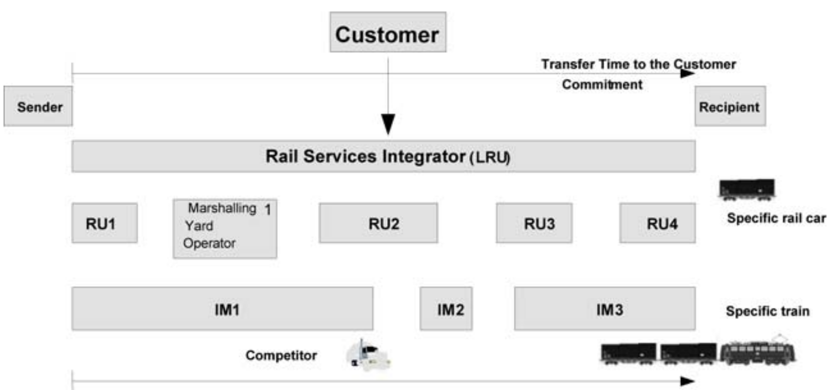


Fig. 1. Area covered by the TAF TSI specification. Source: Own study based on [8].

TSI Telematic Applications include only applications for freight transport services and management of connections with other modes of transport, which means that it focuses on the transport services of the Railway Carriers, as an addition to the operation of trains. The area covered by the specifications is illustrated in Fig. 1.

The transport process model, defined for the purposes of the TAF TSI, implies participation of many market participants in the process and is based on the following assumptions:

- European Railway Sector has been undergoing significant changes associated with the liberalisation of the railway market, open access to infrastructure, purchase/sale contracts, the substitution of RIV and the introduction of the quality charter in freight traffic. Those factors and strong market pressures require significant changes in the processes of cooperation and electronic data exchange to improve the competitiveness against other modes of transport.
- IM and RU compete in the freight traffic network business with other modes of transport (e.g. road vehicles) using simple business processes, supported by IT (information technology) requirements.
- TAF TSI determines boundaries between the cooperation and competitiveness in the European railway sector.
- There are recognisable changes in the European market to improve the customer service upon implementation of the TAF TSI.
- The required flow of messages to the Railway Services Integrator (the leading RU) was determined in TAF TSI. The concept of a leading operator (LRU) refers only to its function performed for the customer in a business process.
- RU can operate on the infrastructures of several IM managers (open access).
- The task of TAF TSI working group (formed of experts from the European railway sector, i.e., RU, IM and rail cars owners) was to define and describe the processes outlined above and related flows of messages.

Model of the transport process in TAF TSI approach is presented in Fig. 2.

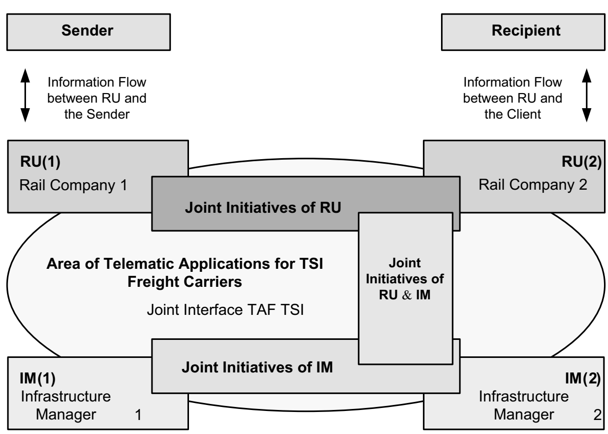


Fig. 2. Model of the transport process in TAF TSI approach. Source: Source: Own study based on [8].

The TAF TSI is focused on providing IT support for business processes in rail freight transport, which may lead to a substantial increase in the quality of transport services. Because of its nature, the TSI shall materially affect the business and operational processes throughout the European rail industry. In addition, a continuous growth in international freight traffic requires a European perspective of information management. Those facts together require defining a coherent trans-European implementation plan for the TSI. It is recommended that such a plan should provide both a vision of what is to be achieved through the implementation of the TSI and that by what means and in what time horizon a transfer should be made from the current situation, characterised by the use of fragmented information systems, to the European system of vast information highway that will provide added value for all actors of rail transport – the infrastructure managers, rail operators, consignors of goods, and finally also for the customers.

In this context, a development of the Strategic European Deployment Plan (SEDP) to organise effective application of TAF TSI specification was ordered. The SEDP defines the target system, which is to be achieved through the implementation of the TSI. The SEDP sets out the way of TAF TSI implementation. It creates a possibility to control the project implementation at the European level. The Strategic European Deployment Plan (SEDP) supports:

- coordination, synchronisation, and prioritisation in the European rail transport;
- allocation of limited resources and budgets to meet business requirements;
- optimisation of the overall costs/benefits;
- protection for shareholders' investments during the implementation phase;
- monitoring the work progress at a company level and at the European one.

Rail Transport Service Information System

The European Commission had decided to grant the EU financial assistance to the Ministry of Transport (now Ministry of Infrastructure) of the Republic of Poland for the implementation of R&D project – "Feasibility Study for Railway Transport Service Information System – SITKol." The project consists of three substantive phases:

- Feasibility Study
- Project Study
- Pilot Implementation

The beneficiary (MT) has appointed the Railway Telecommunication „Telekomunikacja Kolejowa" company for the project implementation. According to the assumptions, the SITKol is to become a system to inform the public about rail transport services and that shall allow, inter alia, for obtaining current information about transportation capacities of rail operators, about decisions (licenses, certificates, etc.) issued, and that will be able to become a tool to monitor traffic. The system is to develop tools to provide equal access of the carriers to the railway transport market in Poland, and to facilitate alignment with the EU legislation for this market. The system should also ensure direct access to information from fixed and mobile terminals and allow for electronic information flow between railway companies, offices, and railway clients.

It is assumed that the exchange of data between participants in the SITKol system will be based on the network of the Railway Telecommunication framework company. The SITKol system is a multi-service platform and its users will be able to enjoy the exchange of data or information through the use of e-mail, websites, WAP, SMS, Contact Centre and telephony services. The data obtained will be made available to authorised users by using commonly applicable protocols such as GSM / GSM-R (SMS, WAP), GPRS, EDGE, UMTS, Wi-Fi and WAN, LAN, WLAN.

The main objectives behind the implementation of the SITKol System are to:

- gain knowledge about the existing information systems related to the SITKol system;
- review the SITKol system architecture;
- verify the demand for information and its impact on the infrastructural costs of the system;
- verify the feasibility and costs of the SITKol project. This system, by direct access to fixed and mobile terminals, should allow for exchange of and sharing information between the Ministry of Infrastructure (MI), the Rail Transport Office (UTK), the rail transport participants, and the direct users of the rail transport.

Basic assumptions of the SITKol system architecture

The system is dedicated to fulfil information needs of the rail transport market participants. The system provides a single point of access to the information related to the passenger and freight transport. Information needs of the rail transport market participants in conjunction with the existing data sources are the basis of the processing and ordination in the SITKol System. The general diagram of the rail transport service information system and the proposed services are illustrated in Fig. 3.

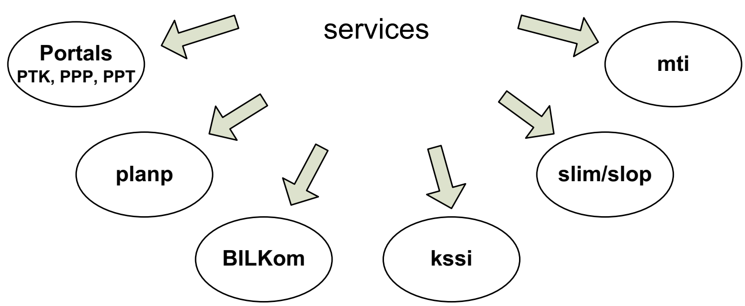


Fig. 3. Diagram and services provided by the rail transport service information system

The SITKol system supports the processing of source data through:

- syndication of information and meta-information about data scattered in various sources,
- classification of the data received in order to make it searchable and sharable from a single point of access,
- data transformations to forms compliant with the European Union requirements – interoperable data exchange,
- feeding the data received to other specialised systems.

The SITKol system supports technical processing of the source data through:

- data downloading via automated interfaces to the source systems,
- data aggregations to meet customer needs,
- sharing of (aggregated and source) data using graphical and machine interfaces,
- using diversified communication channels – commutation of information.

Services performed by the SITKol system are addressed to:

- individuals as recipients of information on the passenger transport.
- business entities as recipients of information on the freight transport.
- carriers as participants in the railway market and the owners of information on the freight and passenger transport.
- managers of various kinds of infrastructure related to railway operations as owners of the information on the basic railway infrastructure.

In the face of growing demand of users for information and the expansion of public resources, the system architecture should be adapted to a continuous and planned development and liable to change, which means that it should be:

- scalable – be expandable by increasing its efficiency, capacity flow, filling capacity without having to replace entire hardware units,
- flexible – have a possibility to use open technology standards to enable easy building of new services into SITKol,
- manageable – able to use stabilised technology standards, and avoiding unproven technologies.

Logical architecture of the rail transport service information system is illustrated in Fig. 4.

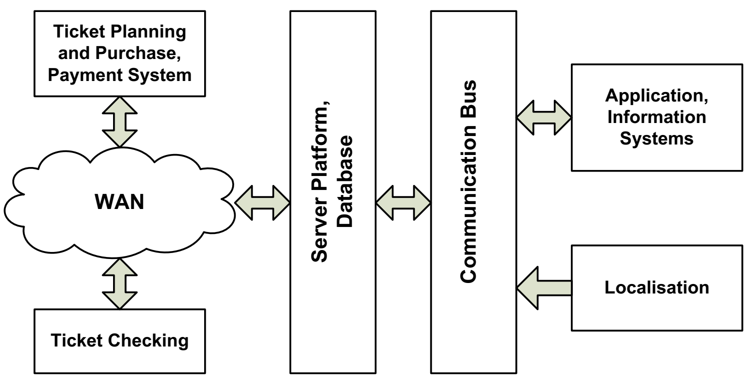


Fig. 4. Logical architecture of the rail transport service information system

5 Conclusions

The data exchange in the transport system can provide greater transparency of operations for all tasks and consequently improve the quality of transport service, which will enable continuous and up-to-date information, improve rail services, and may increase customer satisfaction from the implemented transport service, thus attracting them to use railways whenever having future transportation tasks.

The SITKol system via the operating mode should allow for introduction of any changes in the draft of rail cars, information on delays, cancellations, distractions of rail traffic (e.g. planned and unplanned repairs and repairs of rail tracks), the localisation of trains in real-time, mobile search of connections via a mobile phone, etc.

Implementation of System SITKol may be a basis for a conclusion that at present there are already available telematic tools to create integrated information systems for the rail transport in a comprehensive manner, which will facilitate travel planning, trip fare collection, and travel monitoring. This applies to transport of both passengers and freights. The SITKol System Platform is an open platform and systemically prepared for area and service expansion. Further development of the SITKol System could informatively integrate railway communication with that of the public metropolitan cities – the organizers of the "2012 Event," introduce air and bus transportation to travel planning, or e.g. allow for city-visit planning so to include the best points of interest in a given metropolitan area (the so-called POI).

References

1. Directive 2001/16/EC of the European Parliament and the Council of Europe of 19, on the interoperability of the trans-European conventional rail system - Annex II - Telematic applications (March 2001)
2. Directive 2004/49/EC of the European Parliament and the Council of Europe of 29 April 2004 on safety on the Community's railways (2004)
3. Directive 2007/2/EC of the European Parliament and the Council establishing an Infrastructure for Spatial Information in the European Union, INSPIRE (2007)
4. Communication from the European Communities Commission. reight transport logistics action plan. COM (2007) 607, Brussels (2007)
5. Siergiejczyk, M.: Exploitation Effectiveness of Transport Systems Telematics. Scientific Works of Warsaw University of Technology. Transport Series, vol. (67). OW PW, Warsaw (2009)
6. Siergiejczyk, M., Gago, S.: Convergence of services in IP networks for Railway Companies. In: Materials for 4th Conference of Telecommunications and Computer Science at the Railways. Polish Chamber of Producers of Equipment and Services for the Railways (CD), Szczyrk (2006)
7. A directional strategy for the development of Poland's computerisation until 2013 and prospective forward-looking transformation of the information society by year 2020, Warsaw (June 2005)
8. Wawrzyński, W., et al.: Methods of usage of telematics means in supporting fulfilment of transport tasks. Report from KBN grant 5T12C 066 25, Warsaw (2006)
9. Technical Specifications of Interoperability for Freight Telematic Applications - TAF TSI. Communication from the Commission to the Council of Europe, the European Parliament, European Economic and Social Committee of the Regions - Freight Transport Logistics in Europe – the key to sustainable mobility, Brussels (2006)

Diversity as a Means for Reliability and Safety

Christo Chrsistov[1], Nelly Stoytcheva[2], and Maria Christova[2]

[1] European Polytechnic University,
158 Geo Milev str., 1574 Sofia, Bulgaria
[2] Faculty of Electrification and Telecommunication, University of Transport,
158 Geo Milev str., 1574 Sofia, Bulgaria
{Chchristo,nstoytcheva,maryiahbg}@yahoo.com

Abstract. The diversity as a fault-tolerance and fail-safe method is the issue. The quantity evaluation is used for the depth of the diversity and formulas are derived for the probability of fault detection and dependability function as the diversity. The logic-probabilistic transition method is used.

Keywords: safety, diversity, reliability, fail-safe, fault-tolerance.

1 Diversity and Its Possibilities

1.1 Diversity as a Method

Diversity *is a method of problem solving (mathematical, logical, technical, etc.) in two (A and B) different ways on the same input data.*

A malfunction of hardware or software errors in any of the two decisions, including the specification and algorithm of the task can be found through diversity. From the assumption it results that in the task there are at least two different solutions, which may be perfect (errorless) or have errors. The c**riterion of perfection** *(no errors and faults) is the* ***equivalence*** *(in the particular case – the identity) of the resulting solutions in both results.* In tests of a program error and in the other, the results are different. By comparing the results of logical operation "AND" the (compliance) error "is captured". The comparison may be equivalent or anti equivalent and so based on the other compliance.

The need for distinction between the two solutions stems from the understanding that when the information is handled in different ways, probably the other way (algorithm) to have no errors is valid or if they exist, they are not the same and their influence leads to different results.

1.2 Examples of the Diversity

The equation $ax^2+bx+c=0$ must be solved. There are at least two possible methods: A - classical formula for roots of quadratic equation, and B - formulas in Viet. When you ask the same values of trinomial coefficients, you must obtain the same roots with both methods. If they are not relevant, one of two ways or both were incorrectly programmed.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 308–319, 2010.

It is possible to solve the task twice in the same way (method, program). Then we say about a homogeneous and not diversity decision. In homogeneous digital systems errors cannot be detected, e.g. caused by impulsive disturbances in both channels transform unit to zero or vice versa. If there is an error and both processes are giving the same wrong result, but unknown by comparing.

Let channels are diverse, in particular inversion, where all logical values of one channel are anti equivalent to the other channel. Error (false) transition $1 \to 0$ will change the logical value of 0 only where it is 1. In the other channel it is inverse - 0 and remains 0, even when the disturbance is spread over it. Conversely, false transition $0 \to 1$ creates an error only in this channel, which at this point is 0. In the other channel the value of 1 remains. Instead, *anti-*, equivalent potentials are obtained. When comparing the anti equivalent any error will be detected.

In Fig.1 channel A decides a logical task

$$y = x_1 x_2 \tag{1}$$

and channel B

$$y = \overline{\overline{x_1} \vee \overline{x_2}} = \overline{x_1 x_2} = \overline{y} \tag{2}$$

The control for anti equivalent is carried out continuously under the scheme “a sum module 2”.

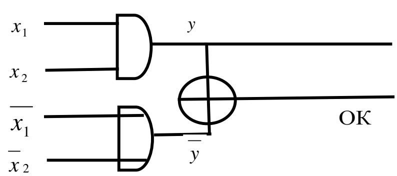


Fig. 1. Example for the diversity structure

In tests of dual-channel processing of the two outputs anti equivalent values are always received and the control output of the signal is „OK”. When in one of them an error appears, both channels receive the same signal and „OK” is removed. The error is detected.

In general, the structure of the system involves dual channel (Fig. 2). After processing and/or transfer of the information the results have to be compared. When a diverse system works perfectly, the comparison has a positive outcome („OK”). But if there is an error in channel A or B the output results will be consistent and „OK” is removed.

Diversity in hardware is achieved by the two decisions of the same task using different pieces of hardware, e.g. devices with the same purpose produced by different companies (Siemens and Thales) or modules operating on different principles (e.g. pir-detector and a microwave detector).

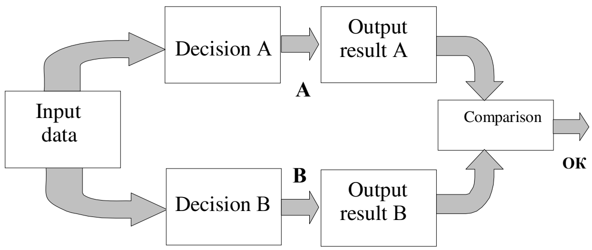


Fig. 2. Structure of the system

Diversity in software is most often achieved through the diversity and independence of the programming teams (developers) solving various versions of the same task. In complex Real Time Systems, such as Signaling Systems, the versions of the two teams are always different. The distinction may appear singly or in combination: the concept, the algorithms in programming languages, representation of data (inversion, reversion or other line of the two records), the source of the program, documentation, etc.

1.3 Faults, Errors and Diversity

The properties of errors must be distinguished from these of faults. While errors occur during operation in different channels (hardware and / or software units) – separate and independent from one another, the errors in the creation of devices (design, documentation, production at the plant, software, etc.) are the same for a lot. So the whole series produced by hundreds and thousands of specimens containing the same errors. If two channel systems be built they are homogeneous and errors remain undetected by comparing their outputs.

The diversity is the most effective means to detect errors. Besides the example in Fig. 1 we may bring many others. If the device of company A does not withstand low temperatures and makes errors, and that of company B is working properly (one gives a result 1, the other - 0), when they are compared the errors will be detected.

Limitations in the ability of diversity will be determined by the probability in both versions to get wrong, but the same results, and hence - unrecognizable by faults comparison. The importance of this likely bad probability depends on two things:

- The likelihood of both programs to prevent the same errors. If developers of the two versions are more independent, the more likely it is to avoid such errors.
- The number of digits w compared to the output vectors (Figure 2). When they are short (e.g. - $w = 1$) the probability to obtain unknown errors is very high. For even distribution of false results it reaches ¼.

1.4 Two Aims of Diversity

Diversity as an approach can be used to achieve two conflicting goals:

- To detect errors and faults;
- To mask the errors and faults.

An error can be identified, a fault – recognized and the execution of erroneous operations may be suspended while eliminating the reasons for fault. In the case of railway signalling systems, the rail traffic stops when an error is identified. Therefore the detection of errors associates with fail-safe systems behaviour and diversity can be regarded as fail-safe (fail-stop) means.

As one of the versions may encounter an error and the other works on these input data, the diversity may have the opposite purpose - to tolerate, to mask, to suppress errors. A diversity system can be composed of the "OR – OR". One of the two channels is a backup and if the other stops working, it is incorporated to work. Whichever of the two channels operates, the system works. To achieve a greater reliability and diversity it is necessary to speak as a means of fault-tolerance. The purpose of this study is to find models that do not identify likelihood of errors and faults when the diversity is used as **fail-safe**, and so the quantitative relationships to influence of diversity on the reliability where it is a **fault-tolerance** means. On this basis to draw conclusions about how to change the influence on the factors so as to achieve a full recognition of errors and better reliability.

1.5 State of the Art

A trial of decision task analysis by identifying the faults diversity placed above was made in [3], [4], [5], and [6]. Functions of the diversity are explained as fault-tolerance means, but no quantitative measure of its "depth". The result remained as a deficit in the diversity science. In particular, there are no analytical relationships, which explicitly identify factors that (together with depth) affect reliability. Only the original publications [1] and [2] sought a quantity measure, but the use of diversity feature to be simultaneously fail-safe and fault-tolerance means and its scientific interpretation is presented here for the first time.

2 Diversity as a Fail-Safe Means

2.1 Quantity Definition of the Diversity Depth

An absolute diversity would mean that:

- errors in both versions are completely independent (not correlated)
- there are no common reasons for faults and the decision of the two ways that leads to the same wrong result.

Let a diversity system has a fault flow with a fault rate of $\lambda(t)$. They are divided into two kinds of faults:

α - Faults: detectable by comparing the output results;

η - Faults: undetectable by comparison, because the same cause errors in the results compared.

Sources of η faults comprise:

- Common source from which the information flows into the diversity system,
- Common components of systems beyond the two channels,
- Sole comparator to compare the results of exit,

- Synchronization of work on both channels,
- Common power supply and more.

In this study an exponential fault distribution may be assumed, where $\lambda(t)$ = const $= \lambda$. This is most often the case in Signaling Systems. The division of faults starts from the assumption that irrespective of any measures taken the diversity is never absolute. The diversity means a difference, but how to measure the difference, the depth of diversity?

The difference can have several perspectives: the degree of independence, decision method, principle, programming language, etc. Based on these differences we can hardly find a quantitative measure. They also do not represent an interest and study. It is interesting, whether through the difference we will be able to find errors. This is the aim to introduce a measure of diversity depth Ω, appropriate with its feature to detect errors:

$$\Omega = \frac{\lambda_\alpha}{\lambda_\alpha + \lambda_\eta} \tag{3}$$

where λ_α, λ_η - fault rates of two fault types.

Since there will be faults of α and η type therefore logically, the two are related conjunctively, then from the theory of reliability it is known that $\lambda_\alpha + \lambda_\eta = \lambda$. If all faults are a common cause faults,

$$\Omega = \frac{\lambda_\alpha}{\lambda_\alpha + \lambda_\eta} = \frac{0}{0 + \lambda} = 0 \cdot$$

For example, if two program versions A and B are equal and have the same errors in both copies of duplicated only program, then the results will be identical, but wrong. The value recognizable in the comparison comes down to zero.

If the two channels are more different, the flow λ_η is smaller, but the flow $\lambda_\alpha = \lambda - \lambda_\eta$ is higher. In the case of $\lambda_\eta = 0$ the diversity is:

$$\Omega = \frac{\lambda_\alpha}{\lambda_\alpha + \lambda_\eta} = \frac{\lambda_\alpha}{\lambda_\alpha + 0} = 1 \tag{4}$$

This achieves the highest recognition of errors and faults.

2.2 Determining the Probability of Fault's Non-identification in the Diversity Computer Systems

In the context of probabilistic logic it can be assumed that the fault would not be identified in two cases:

1. If there is a η fault, which by definition leads to the same erroneous results;
2. If in both channels (α_A and α_B) the α fault exist - and they accidentally cause the same erroneous outcomes.

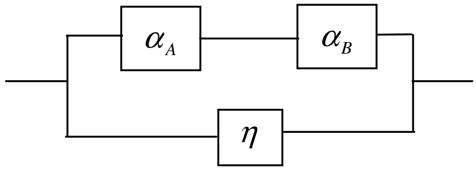


Fig. 3. Logic of fault's non identification

This logic is clearly illustrated in Fig. 3.

Boolean function of fault non identification F_{ni} has the form:

$$F_{ni} = z^1_{\alpha_A} z^1_{\alpha_B} \vee z^1_{\eta} \tag{5}$$

where α_A and α_B are events, which occasionally cause the same erroneous outcomes of the two channels. They represent logical variables $z^{0/1}_{\alpha A}$ and $z^{0/1}_{\alpha B}$ in which 1is a claim that the event happened, and 0 - that never happened.

(5) shows that event F_{ni} can happen when a non-identified η fault happens (logical variable z^1_{η}) and when the two channels is due to random faults unrecognized uniformity of output vectors that cause. Reflected through intersection of two variables - $z^1_{\alpha A}\ z^1_{\alpha B}$.

To be a modeled quantity the influence of the diversity on the identification of faults has to be modeled by a logic-probabilistic transition from (5) to the probability of fault non-identification. After applying the theorem of De Morgan on (5) a non repeatable Boolean function in the underlying "intersection – negation" is obtained:

$$F_{ni} = z^1_{\alpha_A} z^1_{\alpha_B} \vee z^1_{\eta} = \overline{\overline{z^1_{\alpha_A} z^1_{\alpha_B}}.z^0_{\eta}} \tag{6}$$

suitable for replacement with probability values. By substituting out the rules of logic-probability transitions [...], the obtained value is likely to look:

$$Q_{ni}(t) = 1 - \left\{1 - \left[Q_{\alpha_{A\Sigma}}(t) Q_{\alpha_{B\Sigma}}(t)\right]\right\}[1 - Q_{\eta}(t)] \tag{7}$$

where: $Q_{ni}(t)$ is a probability for a non-identified fault, and $Q_{\alpha_{A\Sigma}}(t)$ and $Q_{\alpha_{B\Sigma}}(t)$ are probabilities of independent origination of each other faults, which are in each channel as a result of a chance to obtain the same outcome, but wrong in both output channels A and B; $Q_{\eta}(t)$ - probability of fault occurrence in general components of diversity system, which bring one and the same wrong output signal non-identified by comparison.

In A-and B-treatments the probability due to a fault of obtaining matching errors and illegible by comparing output vectors may occur because of any matching of homonymous vectors pairs. Then the information contained in components $Q_{\alpha_{A\Sigma}}(t)\ Q_{\alpha_{B\Sigma}}(t)$ of (7) could be written as the sum of probabilities:

$$Q_{\alpha_{A\Sigma}}(t) . Q_{\alpha_{B\Sigma}}(t) = \sum_{i=1}^{2^w - 1} Q_{\alpha_{Ai}}(t) Q_{\alpha_{Bi}}(t) \tag{8}$$

where $Q_{\alpha_{ai}}(t)$ and $Q_{\alpha_{bi}}(t)$ are probabilities that the match will be achieved through i-th output vector of the two treatments. One of these is subject to random coincidences functional (working) vector. As we search for a non identification fault, and if it is a fault, this leads to erroneous vectors, the functional vector $i \in \{1,2,... 2^w - 1\}$ should be excluded from considerations.

The summing of probabilities as in (8) is the only possibility if individual events that they model cannot happen simultaneously. It really is – when a pair is generated, another pair of matched vectors may not exist at this time. If all error failed vectors in a given diversity unit (program, computer controller) are equally likely, then:

$$Q_{\alpha_{A\Sigma}}(t) \cdot Q_{\alpha_{B\Sigma}}(t) = (2^w - 1)\, Q_{\alpha_{Ai}}(t) Q_{\alpha_{Bi}}(t) \tag{9}$$

Where $Q_{\alpha A}(t)$ and $Q_{\alpha B}(t)$ are likely to get any fault vector of A or B treatment, respectively. If the distribution is with equal probability, the following can be written for (9):

$$Q_{\alpha_{A\Sigma}}(t) \cdot Q_{\alpha_{B\Sigma}}(t) = (2^w - 1)\, \frac{Q_{\alpha_A}(t) Q_{\alpha_B}(t)}{2^{2w}} \tag{10}$$

The equation reporting that at a given time from possible errors there is one that matches the functional vector. It is likely to occur:

$$Q_{\phi}(t) = \frac{Q_{\alpha_A}(t)}{2^w} \frac{Q_{\alpha_B}(t)}{2^w} \tag{11}$$

Then

$$Q_{ni}(t) = 1 - \left\{1 - (2^w - 1)\frac{Q_{\alpha A}(t) Q_{\alpha B}(t)}{2^{2w}}\right\}\left(1 - Q_{\eta}(t)\right) \tag{12}$$

When the distribution is exponential, the faults' rates $\lambda_\alpha \, and \, \lambda_\eta$ are independent of time values. The use of known relationship between the rate and probability of fault $Q(t) = 1 - e^{-\lambda t}$ and replacing it in (12) gives:

$$Q_{ni}(t) = 1 - \left\{1 - (2^w - 1)\left[\frac{\left(1 - e^{-\lambda_{\alpha A} t}\right)\left(1 - e^{-\lambda_{\alpha B} t}\right)}{2^w}\right]^2\right\} e^{-\lambda_\eta t} \tag{13}$$

If we have an equal probability of faults in both channels $Q_{\alpha_a}(t) = Q_{\alpha_b}(t)$ is obtained:

$$Q_{ni}(t) = 1 - \left\{1 - (2^w - 1)\left[\frac{\left(1 - e^{-\lambda_\alpha t}\right)}{2^w}\right]^2\right\} e^{-\lambda_\eta t} \tag{14}$$

Instead of fault rates λ_α and λ_η the depth of diversity Ω may be introduced from(4): $\lambda_\alpha = \Omega\lambda$ and $\lambda_\eta = (1 - \Omega)\lambda$. After substitution (14) will become:

$$Q_{ni}(t) = 1 - \left\{ 1 - (2^w - 1)\left[\frac{\left(1 - e^{-\Omega\lambda t}\right)}{2^w} \right]^2 \right\} e^{(\Omega-1)t} \tag{15}$$

Obviously if diversity is maximum and the likelihood and $\Omega = 1$ then non-identification is minimal:

$$Q_{ni\min}(t) = 1 - \left\{ 1 - \left(2^w - 1\right)\left[\frac{1 - e^{-\lambda t}}{2^w} \right]^2 \right\} = (2^w - 1)\left[\frac{1 - e^{-\lambda t}}{2^w} \right]^2 \tag{16}$$

It seems that after a time long enough the probability for fault comes to 1. It appears that after a time long enough, when the probability of fault approaches one:

$$Q_{\substack{ni \\ t \to \infty}}(t) = \frac{1}{2^w}\left(\frac{2^w - 1}{2^w} \right) \tag{17}$$

and this is the maximum of the non-identification likelihood of the absolute diversity $\Omega = 1$.

If the number of digits of *w* of the output compared vector is higher, they are less likely to coincide by chance after a fault. The worst case would be for one digit. If $w = 1$ in each of the two channels, whose outputs are compared, there is one binary output. When $\Omega = 1$, the minimum probability of fault non-identification would have a maximum value:

$$Q_{ni\min_{\max}}(t) = \left[\frac{1 - e^{-\lambda t}}{2} \right]^2 = \frac{1}{4} Q(t)^2 \tag{18}$$

This means that if two diversity channels have the same fault rates λ and their diversity is absolute, the probability of fault non-identification of a two-channel system with one physical output would be ¼ of the square of the fault probability in each of independent channels. This could be expected, because of the channels independence and we have equal fault probabilities of this issue that can have 0 or 1 value, a wrong match of the same results will be possible in 25% of cases.

When there is no diversity ($\Omega = 0$), the probability of a non-identification (15) is greatest and is equal to that of a single channel system, whose fault rate is λ: $Q_{ni_{\max}}(t) = 1 - e^{-\lambda t}$.

Let us explicitly clarify the premise of these conclusions: by the diversity depth Ω fault rates λ are distributed between the general components of the system on the one hand, and between both individual and independent channels A and B - on the other.

The analysis shows that the probability of non-identification of fault $Q_{ni}(t)$ depends on the diversity depth Ω of the system, on the number of digits *w*, which present the outcome of treatment, on the total fault rates in the system λ and on the system time t. How this probability will vary can be determined if the relevant calculations are made using established formulas. But they also show that in other equal conditions the diversity depth and the number of digits of the output vectors are crucial for the identification of faults.

3 Diversity as a Means of Fault-Tolerance

The diversity may have the opposite purpose - to tolerate, to mask, to suppress errors, rather than to identify them. Where one program has activated an error, so that the program is incapable, in the alternative treatment program it is not very likely to have an error and to obtain a true solution. One of the two programs can block, but the other displays a correct result. The diversity can be used to create a fault-tolerant system that tolerates a large part of faults and faults.

There are two implementation techniques:

- The error was found to be switched to the second program. In this case we speak about **reservation through substitution;**
- Both options (programs, channels) are always used in solving any problem. Then the talk turned to a **permanent reserve**. The two channels are equal, no basic and reserve channels.

Let the rate of independent faults in both channels is λ_α. In the event of a fault, the system cannot work because the other working channel is independent from failing. In terms of efficiency scheme the reliability can be modeled as shown in Fig. 4.

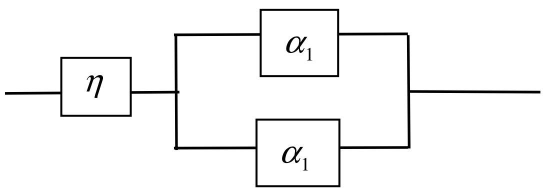


Fig. 4. Reliability model

The logic of such a scheme is as follows: the system will work if there is no common case fault (η-fault) and at least one channel has no α - fault. A Boolean function of the efficiency of the scheme has the form:

$$F_p = z_\eta^0 \left(z_{\alpha 1}^0 \vee z_{\alpha 2}^0 \right) \tag{19}$$

Since the Boolean function is not repeatable, it can be used for transition to a probabilistic function. But for this aim it should be in the underlying "intersection-negation". This form is obtained through the following transformation based on Morgan:

$$F_p = z_\eta^0 \left(z_{\alpha 1}^0 \vee z_{\alpha 2}^0 \right) = z_\eta^0 \cdot \overline{z_{\alpha 1}^1 z_{\alpha 2}^1} \tag{20}$$

By applying rules to transfer from logic to probability functions eq. (20) is a formula for the probability of reliable operation of a two-channel system:

$$P_s(t) = P_\eta(t)\left[1 - Q_{\alpha 1}(t).Q_{\alpha 2}(t)\right] \tag{21}$$

where $Q_\alpha(t)$ is a probability to get a local fault in the corresponding channel, and the likelihood of disability due to the absence of common causes. Assuming exponential distribution of time, the formula acquires the form:

$$P_s(t) \underset{\lambda=const}{=} e^{-\lambda_\eta t}\left[1-(1-e^{-\lambda_{\alpha_1}t})(1-e^{-\lambda_{\alpha 2}t})\right] \tag{22}$$

In the particular case where both channels have the same rate of their own faults $\lambda_{\alpha 1}=\lambda_{\alpha 2}$:

$$P_s(t) \underset{\lambda=const}{=} e^{-\lambda_\eta t}\left[1-(1-e^{-\lambda_\alpha t})^2\right] \tag{23}$$

Since $\lambda_\eta=(1-\Omega)\lambda$, and $\lambda_\alpha=\Omega\lambda$, after substitution in (23) the probability for the reliability function is:

$$P_s(t)=e^{-(1-\Omega)\lambda t}\left[1-(1-e^{-\Omega\lambda t})^2\right] \tag{24}$$

This formula determines the reliability of dual-channel diversity system, which has common and private reasons for faults in relation Ω and where local channels in both reasons have the same fault rates.

Although private, this case allows us to do some research important to reliability. But now it appears that when Ω = 1 (where the best diversity) it is equivalent to the system in parallel two channels system:

$$P_s(t)=1-(1-e^{-\lambda t})^2 \tag{25}$$

When Ω = 0, i.e. both channels are identical, or the same channel is used twice (homogeneous), the reliability dropped sharply:

$$P_s(t)=e^{-\lambda t} \tag{26}$$

Let ***n*** channels are numerous and have different rates of local faults. A Boolean function has the form:

$$F_p = z_\eta^0\left(z_{\alpha 1}^0 \vee z_{\alpha 2}^0 \vee \ldots\ldots. \vee z_{\alpha n}^0\right) \tag{27}$$

After De Morgan transformers:

$$F_p = z_\eta^0\, \overline{z_{\alpha 1}^0 z_{\alpha 2}^0 \ldots\ldots. z_{\alpha n}^0} \tag{28}$$

The formula for the probability of reliability operation of the system of ***n*** channels is obtained after a similar logic-probabilistic transition:

$$P_s(t)=R_\eta(t)\left[1-Q_{\alpha 1}(t).Q_{\alpha 2}(t)....Q_{\alpha n}(t)\right] \tag{29}$$

$$P_s(t)=e^{-\lambda_\eta t}\left[1-\prod_{i=1}^{n} Q_{\alpha i}(t)\right] \tag{30}$$

If the fault rate is exponential:

$$P_s(t) \underset{\lambda = const}{=} e^{-\lambda_\eta t}\left[1 - \prod_{i=1}^{n} (1 - e^{-\lambda_{\alpha i} t})\right] \tag{31}$$

In a private case when channels have equal fault rates $\lambda_{\alpha 1} = \lambda_{\alpha 2} = \ldots\lambda_{\alpha i} = \ldots\lambda_{\alpha n}$, and when we account that $\lambda_\mu = (1 - \Omega)\lambda$, and $\lambda_\sigma = \Omega\lambda$, the reliability is:

$$P_s(t) \underset{\lambda=const}{=} e^{-(1-\Omega)\lambda t}\left[1 - (1 - e^{-\Omega\lambda t})^n\right] \tag{32}$$

It is evident that at the amendment of the diversity depth from zero to 1 the system reliability increases as the transition from the coherent (with reliability $e^{-\lambda t}$) in parallel system reliability $\left[1-(1-e^{-\Omega\lambda t})^n\right]$.

To establish the reliability diagram from the diversity depth the calculations were made acc. to the above formulas (32) for a dual-channel case, which are graphically illustrated in Fig. 5. The figure shows that with increasing diversity depth the reliability improves significantly. For example, when λt = 1 and depth Ω = 0.8 the reliability function increased from 0.36 to 0.57.

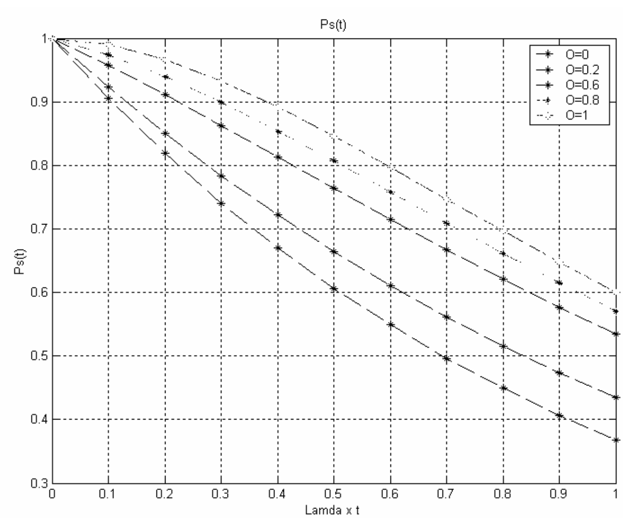


Fig. 5. The reliability diagram

The following important conclusions may be drawn from both models and the research carried out:

1. The deeper is the diversity of channels, the much greater is the reliability. By amending the diversity depth from 0 to 1 the reliability is amended as when a coherent transition to a parallel system reliability.

2. To determine the factors that determine the diversity depth it is necessary to examine the specific scheme for the case by looking for general and local causes of faults and their rates.

Acknowledgements. This paper is supported by the Bulgarian Ministry of Education under the project "Creation of a 'Research and Development Centre' for embedded and mobile devices"

References

1. Christiov, C., Popov, G.: Modeling Diversity as Fault-tolerance Tool. In: International Conference Automatics and Informatics 2009, Sofia 29.09 – 4.10.2009, pp. 127–131 (2009)
2. Popov, G.: Modelling Diversity as a Method of Detecting Faults in non Recovery Computer Systems. Information Technologies and Control 2, 15–19 (2005)
3. Harrison, D.: The use of Computers in safety critical Applications. In: HSC 1998 (1998)
4. Strigini, L.: Diversity: Directions for Research. Centre for Software Reliability. City University, London (2007)
5. Fantechi, S.G., Lombardi G.: Experimenting with Diversity in the Formal Development of Railway Signalling Systems, `http://fmt.isti.cnr.it`
6. Medikonda, S., Panchumarthy, P.S.: An Approach to Modeling Software Safety in Safe-Critical Systems. Journal of Computers Science (5)4 (2009)

One and a Half Year of Participation in the EDCN Project – Summary

Oleg Antemijczuk, Krzysztof A. Cyran, and Eugeniusz Wróbel

Institute of Informatics, Civil Aviation Personnel Education Centre of Central and Eastern Europe, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland
{oleg.antemijczuk,krzysztof.cyran,eugeniusz.wrobel}@polsl.pl

Abstract. The paper describes the results of an engagement in the international cooperation within the EDCN network and it draws conclusions based on experience in monitoring of the EGNOS system. The experience gained by the authors in the EDCN project course is expected to have a serious impact on the research and development dedicated for enhancement of the reliability of the EDCN network. Therefore, in the paper not only are the current results of cooperation presented, but also the challenging issues, which are to be resolved in the future based on the conclusions drawn on the experience gained.

Keywords: EGNOS, EDCN, GNSS, GPS, GALILEO, GLONASS, SBAS, PolaRx-3.

1 Introduction

The development of the European Geostationary Navigation Overlay Service (EGNOS) reached the point at which the European Space Agency could announce the beginning of operation of the system. It happened on September 1st, 2009. The EGNOS 'Open Service' is now available. This is a milestone for the project, since its primary service is now available to all users equipped with EGNOS-compatible receivers. Most mass-market satellite navigation receivers being sold today are ready for EGNOS.

However, the system working in the operational phase needs parallel development of the user segment. Moreover, the introduction of global navigation satellite system (GNSS) for civil aviation requires development of a number of tools and analytical methods for monitoring its safe and continuous work.

The accuracy of global positioning system (GPS) is affected by several physical factors. They include: state of the ionosphere, clocks accuracy of the GPS satellites constellation, precision of computing systems, etc. Increased precision positioning can be provided by additional ground-based measurements supporting GNSS satellites. One of them - the EGNOS system which is based on the network of ground base stations and geostationary satellites transmitting data corrections to the GPS receivers, allows the significant increase of the accuracy using so called differential positioning method. However, both GNSS and EGNOS systems require continuous monitoring.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 320–326, 2010.

For this purpose, the specialized network has been created based on EGNOS Data Connection Network (EDCN) ground stations monitoring the availability and quality of GPS positioning. This system is still under development by the accession of new members from different European Union countries and improvement of diagnostic tools [3].

The EDCN project, described in the preceding paper [1], started in 2001 and it was closed at the end of 2009. Today preparations are underway for the resumption of research in a new project sponsored by EUROCONTROL – the EDCN2. Coordinating a new project is done by the company Pildo Labs from Barcelona. The measurement data collected from the GNSS receiver network are sent to the new FTP server maintained by Pildo. EUROCONTROL organization is currently under negotiation with the European Commission to launch new funds for the EDCN2 project. This project will be extended to support the Galileo satellite constellation and for research on disturbance propagation of satellite signals in the ionosphere of the Earth caused by solar activity. The following article is a summary of our participation from May 2009 to the present.

2 Performance

During more than one year of our participation in the EDCN project, hardware and software installed in the Institute of Informatics process the measurement data gathered from the receiver Septentrio PolaRx-3. The main system server collects continuously all available data from the GPS, SBAS and GLONAS satellite constellations, for the analysis with 1 Hz frequency 24 / 7 days. The collected binary data occupy 113 GB of the disk space on the 1 TB disk array RAID-5. Installed Septentrio communication software collects measurements data from the PolaRx-3 receiver via TCP/IP network and writes it in binary SBF format [5] to the Windows 2008 Server directory [6]. PEGASUS Software ver. 4.5 supplied by the EUROCONTROL organization converts binary data and generates daily reports, which are then sent to a Pildo FTP server and EUROCONTROL research center servers. The software focuses on the analysis of signals transmitted by the SBAS satellites to investigate the continuity, quality and accuracy of the position amendments sent to the GPS receiving antenna.

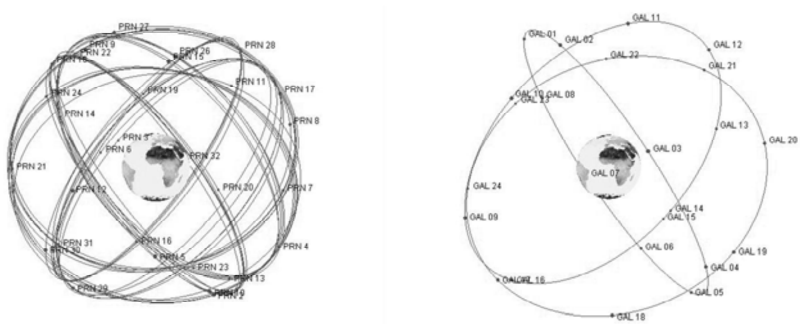


Fig. 1. GNSS satellite constellations – Navstar GPS, Galileo

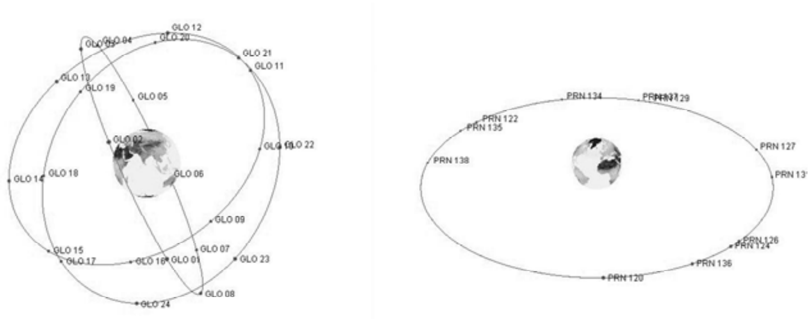


Fig. 2. GNSS satellite constellations –GLONASS and SBAS (EGNOS)

At the same time, the PolaRx-3 receiver software implements multichannel reception of all visible GNSS satellites in the sky [4] (Fig. 1 and Fig. 2).

During the EDCN project (1.5 year), an automatic data transfer was interrupted on several occasions. Breaks in data collection and transmission were caused by the ongoing network infrastructure maintenance, server maintenance data collection system and power modernizing inside the Institute building. Since the master data collecting server is visible on the Internet and is available at the public IP number there is an urgent need to update the system software for safety reasons. Unfortunately, operating system software updates for Windows 2008 Server are very often and require server retention and its reboots. Life activity chart available on the Pildo website for the EDCN project is presented in Fig. 3.

Global Accuracy Performance Summary			
Horizontal Domain			
Mean	Standard Deviation	Maximum Error	HPE 95%
0.3 m	0.21 m	5.8 m	0.8 m
Vertical Domain			
Mean	Standard Deviation	Maximum Error	VPE 95%
0.58 m	0.42 m	7.3 m	1.4 m

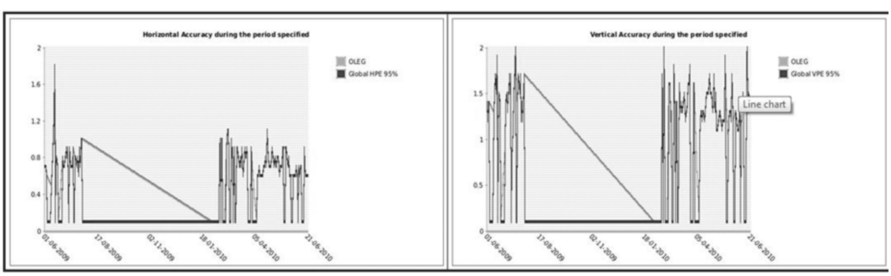


Fig. 3. ZMITAC Server global Accuracy Performance. Source: [6].

Visible break on the graph in the data sending process (horizontal line) is the result of moving measurements and reports database from old EDCN to new EDCN2 server prepared to be launched in a new project [4].

In the course of the EDCN project, on September 1st 2009, the EGNOS system has been forwarded to the general public. This fact was recorded by our software, which generates reports to the EUROCONTROL. Fig. 4 shows a horizontal measuring deviations chart of the GPS position of the EGNOS satellite PRN126 (IOR-W) during the test period from selected two days 15/06/2009 and 24/06/2009. In this figure, a scattering of position accuracy can be observed.

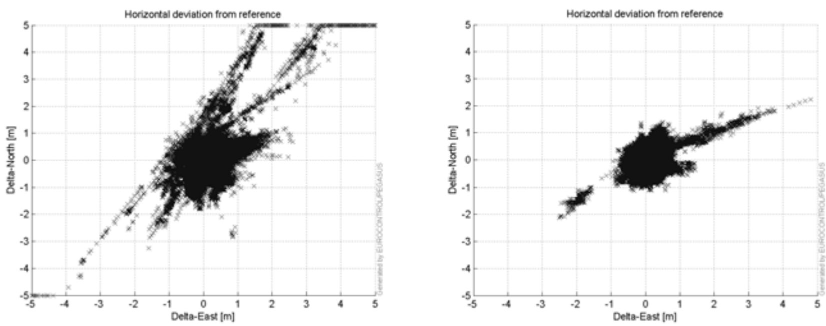


Fig. 4. PRN126 Scatter plot of horizontal deviation from reference position during test period of EGNOS days: 15/06/2009 and 24/06/2009

In Fig. 5 there can be observed the same horizontal deviation graph of PRN126 satellite (IOR-W) after the introduction into service the EGNOS system on 01/10/2009. The diagrams show the day 26/05/2010 and 05/06/2010 situations. A significant increase in the GPS accuracy positioning is visible.

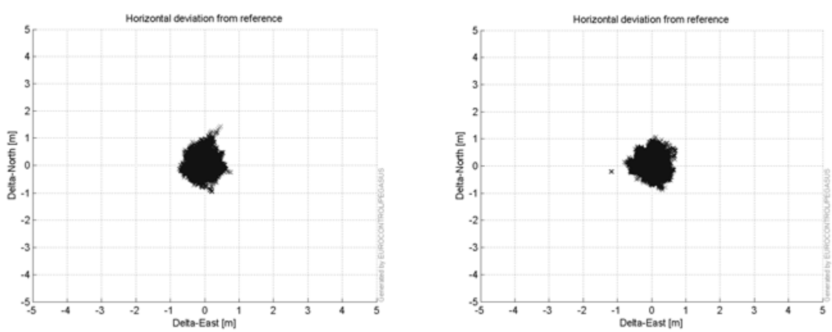


Fig. 5. PRN 126 Scatter plot of horizontal deviation from reference position after operations start of EGNOS days: 26/05/2010 and 05/06/2010

3 Data Discontinuities

During the EDCN project research, the breaks were observed in some SBAS satellite receiving signals needed to determine the exact position of the GPS receiver. In our geographic conditions, in southern Poland, there are visible three satellites - PRN120, PRN124 and PRN126 of the SBAS constellation. Radio signals of these three satellites were received differently within the duration of the project. Fig. 6 shows the availability of the satellite PRN124 signal, which can be compared with the availability of satellite PRN126 signal shown in Fig. 7.

The comparison clearly shows that the availability of PRN126 satellite was worse than that of PRN124 and PRN120. This effect is the subject of the advanced research. The comparison of PRN120, PRN124 and PRN126 satellite visibility from our location, to the PANSA Warsaw and Cracow receivers data received during similar period of time, has revealed that this is a peculiar feature that occurs only in Gliwice. The reason for the absence of signal reception is not the failure of the satellite PRN126, but rather a signal distortion received in our surroundings. Geostationary satellite PRN126 (INMARSAT IND-W) is visible in Gliwice at position 172° E and elevation 30°. At a distance of 3.38 km in the direction of 146.33° E, as shown in Fig. 8, is a mineral wool factory whose smoke from a smoke-stack may, in certain weather conditions, interfere with the reception of the PRN126 satellite signal. It is being explored.

Initial Date: 04-02-2010
Final Date: 21-06-2010
Selected PRN: 124
Selected stations: OLEG

Station Performances				
Station	**Site**		**Location**	**PRN**
OLEG	ZMiTAC		Poland	124
Start	**End**	**SBAS availability**	**Number of Outliers**	**Number of Integrity events**
2010-02-04	2010-06-21	50.2%	0	0

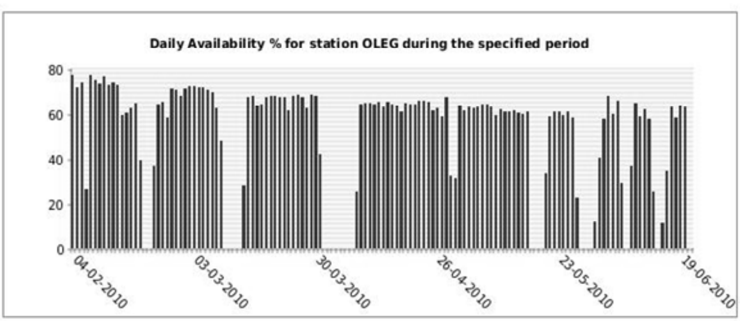


Fig. 6. PRN124 daily availability from 04/02/2010 till 21/06/2010

Initial Date: 04-02-2010
Final Date: 21-06-2010
Selected PRN: 126
Selected stations: OLEG

Station Performances					
Station	**Site**		**Location**	**PRN**	
OLEG	ZMiTAC		Poland	126	
Start	**End**	**SBAS availability**	**Number of Outliers**	**Number of Integrity events**	
2010-02-04	2010-06-21	54.53%	0	0	

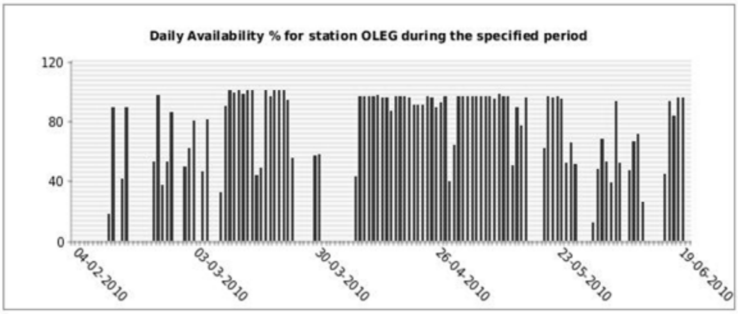


Fig. 7. PRN126 daily availability from 04/02/2010 till 21/06/2010

Fig. 8. Google Earth - EDCN antenna and factory placement

4 Software Development

In our Institute, during the research work there has been created the recording measurement data software for the PolaRx-3 receiver. This software stores all available GNSS and other technical data to an SQL database to provide satellite data in the form of more convenient for further analysis than the data stored originally in SBF binary format. For the purpose of 3D visualization we developed client software presented in Fig. 9. This software graphically shows the position of the GPS antenna perceived by the PolaRx-3 receiver. The figure shows the scattering position within the measuring day (86,400 samples during 24 hours).

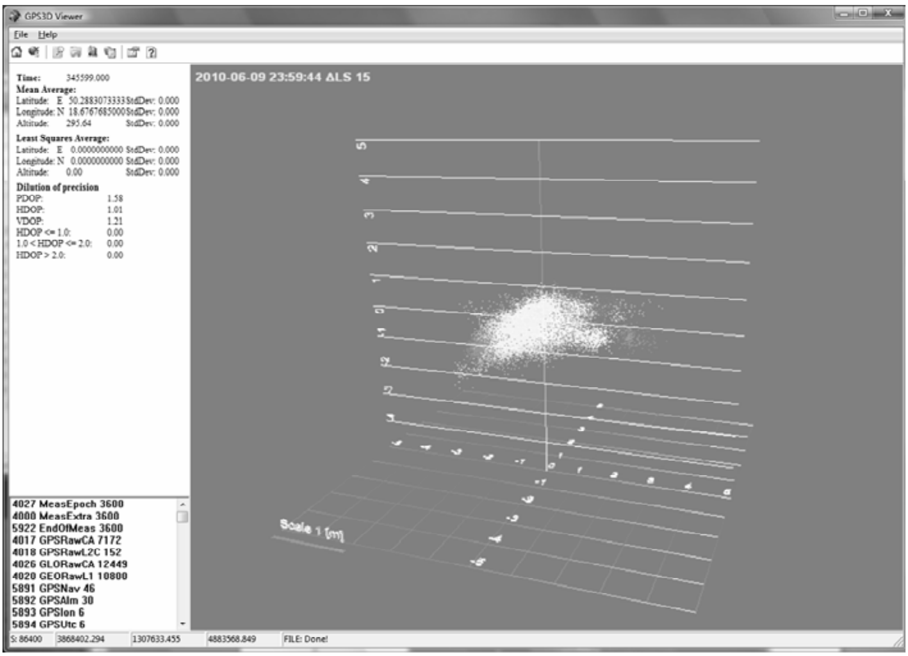


Fig. 9. GPS3D client application view

References

1. Antemijczuk, O., Cyran, K.A., Wróbel, E.: Application of Polarx3 receiver in EGNOS data collection network. Archives of Transport System Telematics 3, 31–37 (2009)
2. ESA official web site, `http://www.esa.int/esaNA/index.html`
3. EUROCONTROL EGNOS Data Collection Network, `http://edcn2.pildo.com/home/`
4. EUROCONTROL Navigation Domain, `http://www.ecacnav.com/NSG_and_Task_Forces/EGNOS_Data_Collection_Network`
5. Septentrio official web site, `http://www.septentrio.com`
6. `http://edcn2.pildo.com/home/`

Application of Wavelets and Fuzzy Sets to the Detection of Head-Checking Defects in Railway Rails

Piotr Lesiak and Piotr Bojarczak

Technical University in Radom, Faculty of Transport and Electrical Engineering,
29 Malczewskiego St., 26-600 Radom, Poland
p.lesiak@pr.radom.pl, bojarczpp@wp.pl

Abstract. Head checking defects pose a major threat to the safety of railway traffic. Because of their location and small dimensions they cannot be detected with the use of traditional ultrasonic methods. An alternative method is an optical method using a laser diode as the transmitter and a PIN diode as the receiver. Head checking defects cause the dispersion of laser rays what in turn causes the generation of undetermined signal containing the information about defects. Because of this the authors dealt with algorithms used to extract and classify these defects. Wavelet transforms and fuzzy sets have been proposed. The efficiency of these methods has been compared.

Keywords: railway defects, optoelectronics, wavelets, fuzzy sets.

1 Introduction

This paper presents an attempt to apply the scattering method to the detection of railway defects [8], [10], [14] and [17]. This method exploits the laser light scattering from rough surfaces. Railway defects investigated belong to the group of contact-stress defects arising during natural exploitation of railway rails [3], [4]. This new group of defects has been known since the last decade of the 20th century. They arise from plastic deformation of the rail surface, and have the shape of scratch, crack or indentation. They are caused by a significant stress occurring between the wheel and the rail. Their typical representative is a head-checking defect [3]. They occur more or less regularly in the shape of small cracks at a distance of 0.5 – 10 mm and at an angle of 10° to 15° to the rail cross-section. These small defects repeating periodically and reaching the size of few millimeters could pose a major threat to the safety of railway traffic [3] – they trigger multi-point breaking of the rail.

Because of their position and small dimensions they cannot be detected with the use of a traditional ultrasonic method [3], [4]. An alternative method uses a laser diode as the transmitter and a PIN photo-diode as the receiver. The assessment of the phenomenon arising during laser rays' scattering on these defects is the basis for the decision making concerning the usefulness of the rail in further exploitation. In order to conduct the experiment a fully computerized laboratory system was built [7].

The right classification of obtained results is also a crucial problem [5], [6]. Thus the extraction of information concerning head-checking defects requires further processing of measuring results obtained from the scattering method. The system designed

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 327–334, 2010.

has been equipped with the following processing blocks: DFT - Discrete Fourier Transform and DWT – Discrete Wavelet Transform [2], [11], Fuzzy Sets [1], [9], [12], [13], [15] and [16] and the Fractal block determining signals' self-similarity features on the basis of circle fractal dimension. This paper presents only algorithms concerning fuzzy sets and a wavelet transform. These algorithms are supposed to eliminate the noise from the signal and extract this part of the signal corresponding to a dangerous defect.

2 Fuzzy Sets in Extraction of Head-Checking Defects

Fig. 1a presents a representative waveform of signal for laser light scattering for a chosen part of rail having head-checking defects. It is created from voltage samples received by a PIN photo-diode and interpolated for better visualization. In this method, at the beginning of measuring process the system in place of the rail head (without defects) is calibrated to obtain a maximal signal value. The maximal signal value is equal to 10V. Voltages from the range of 3.5 - 10V have been assigned to the Noise Set (*N*). Voltages from the range of 0-2.5V have been assigned to the Damage Set (*D*). Voltages from the range of 1.5 - 4.5V with the maximum at 3V have been assigned to the Small Damage Set (*S*). For example, fuzzy set of defect *D* can be presented as:

$$D = \{(u, \mu_D(u)) : u \in U, \mu_D(u) \in [0,1]\} \tag{1}$$

where: $\mu_D : U \to [0,1]$ is a membership function determining the degree of membership of signal samples in set *D* [12], [15] and [16] .

Therefore we obtain the following relationship for consequent samples:

$$D = \mu_D(u_1)|u_1 + \mu_D(u_2)|u_2 + ...\mu_D(u_k)|u_k = \sum_{i=1}^{k} \mu_{Di}|u_i \tag{2}$$

Membership function has been chosen arbitrary. The system should assure a large possibility of choice. Fig. 1d presents a possibility of choosing among different types of membership functions available in the list box. If it is necessary, these functions could be enlarged with the use of higher order polynomials or concatenation of weak monotonic functions [1], [9].

In classification of head-checking defects the membership function of type *Z* has been used. The function of type *s* has been used to reduce the noise. Gaussian function has been used for the Small Damage Set (*S*) – Fig. 1c. Therefore membership functions corresponding to previously assumed voltage ranges could be presented in the following manner:

$$z_D = \begin{cases} 1 & for \quad u \le 1{,}5\,\mathrm{V} \\ 1 - 2(u - 1{,}5)^2 & for \quad 1{,}5\,\mathrm{V} < u < 2\,\mathrm{V} \\ 2(u - 2{,}5)^2 & for \quad 2\,\mathrm{V} < u < 2{,}5\,\mathrm{V} \\ 0 & for \quad u \ge 2{,}5\,\mathrm{V} \end{cases} \tag{3}$$

The laser light scattering of occurs both on thin cracks of pitch surface of the rail head (abrupt decrease of voltage received by a PIN photo-diode) and on long term geometric deformations. It is apparent in the initial part of samples, thus it is important to use an additional amplitude detection Fig. 1b and to choose optimal parameters for membership functions Fig. 1d. In the case of obtaining insufficient results, before fuzzification process signal preprocessing could be applied. It can be realized with the use of remaining blocks of the system.

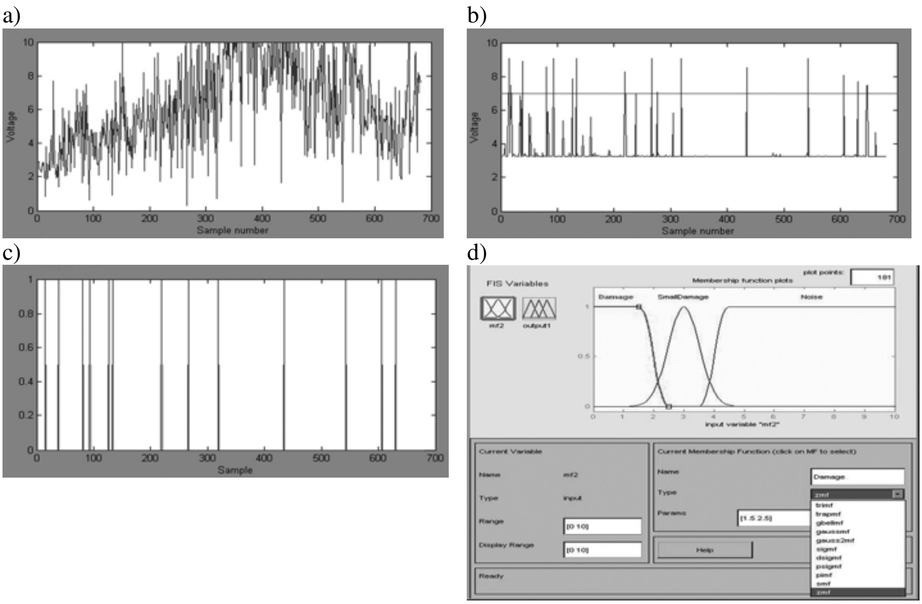
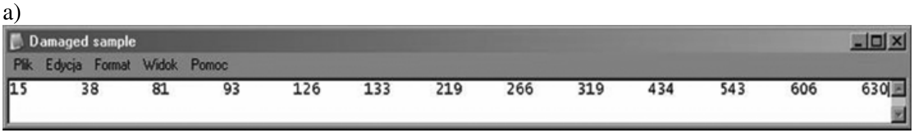


Fig. 1. Software for defects classification using fuzzy logic algorithms, a) registered head checking defects response signal, b) the signal after fuzzyfication of definite detection threshold level, c) classification after using a threshold function, d) a window for signal classification fuzzy function design

a)



b)

1|0,83 1|1,40 0,97|1,63 1|0,61 0,82|1,80 1|0,89 0,91|1,71 1|0,31 1|0,67 0,96|1,64 1|0,46 0,86|1,76 0,78|1,83

Fig. 2. A fuzzy set of samples corresponding to head checking defects: a) text file with numbers of samples qualified as defects (in the frame), b) recording of samples numbers in the notation compatible with formula (2)

Fig. 2a presents a fuzzy set of number of samples assigned to head-checking defects. Elements of this set are placed in the row in Fig. 2b – on the right side (samples expressed in Volts). The kernel of fuzzy set is a crisp set whose elements are samples completely belonging to the set of defects D, what can be written as *ker*(D)={0.83 1.40 0.61 0.89 0.31 0.67 0.46}. In practice they are assigned to dangerous defects.

The quality of decision is usually given as a percentage contribution of right decision for the given set of samples. It can be assessed with the help of a Root Mean Squared Error RMSE [13]:

$$RMSE = \sqrt{\frac{1}{k}\sum_{i=1}^{k}\left|[\mu_{D_i} - d_i]\right|} \tag{4}$$

where: μ_{D_i} is the value of membership function for sample i and d_i is a desired value.

For the example in Fig. 2, the RMSE equal to 0.23 was calculated on the assumption of d_i =1, what is correct for *ker*(D). It means that the classifier based on fuzzy sets correctly classified most of head-checking defects occurring in the rail as dangerous flaws.

3 Wavelet Transform in Extraction of Head-Checking Defects

Thanks to unique features of a wavelet transform, it is commonly used to reduce the noise in the signals analysis. The wavelet transform decomposes the analyzed function into finite lasting components $\Psi(t)$ called wavelets [2], [11]. The continuous-time wavelet transform (CWT) of the function f(t) generates CWT coefficients $W(a,b)$ according to the relation:

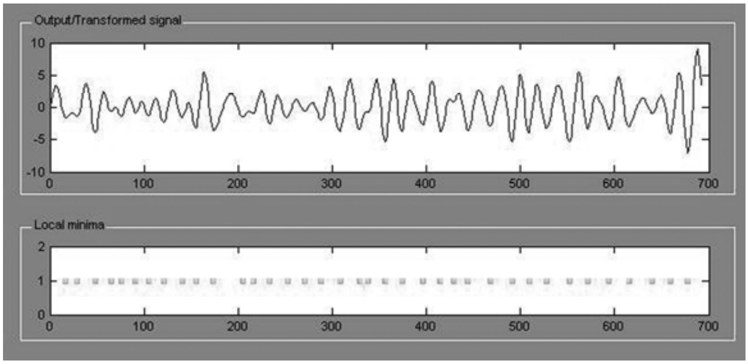
$$W(a,b) = \int_{-\infty}^{\infty} f(t) * \Psi^{*}{}_{a,b}(t)dt \tag{5}$$

where: $\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right)$ is a wavelet function of time scale (dilation) equal a and of time shift (translation) described by b, and $\Psi^{*}{}_{a,b}$ denotes a complex conjugation. For a=1 and b=0 $\Psi_{1,0}(t) = \Psi(t)$ is called the mother wavelet. The wavelet transform is reversible and the original function can be reconstructed on the basis of values of $W(a,b)$ coefficients [2], [11].

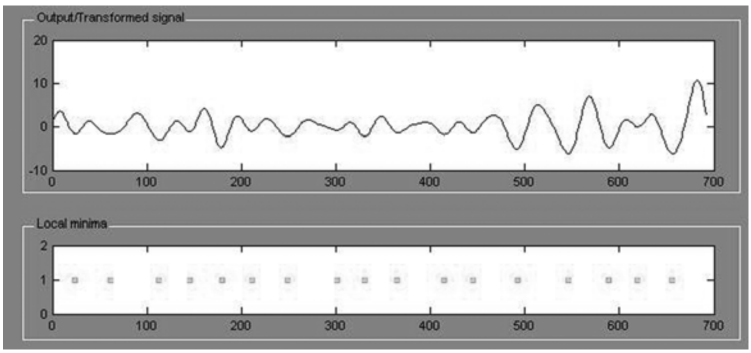
It turns out that it is not necessary to take into account all possible variables a and b. Instead, it is possible to choose a finite set of their values satisfying the following conditions: $a = 2^k$, $b = 2^k n$: where k and n are integers. In this case the original function f(t) can be represented as a superposition of dilated and translated wavelets with the weights $d(k,n)$ denoting the discrete wavelet transform (DWT) coefficients:

$$f(t) = \sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} d(k,n)2^{-k/2}\Psi(2^{-k}t - n) \tag{6}$$

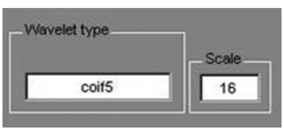
a)



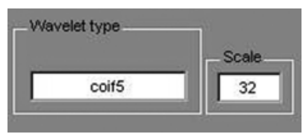
d)



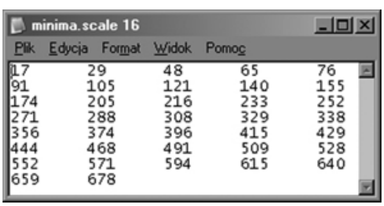
b)



e)



c)



f)

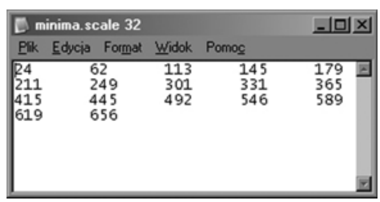


Fig. 3. Realization of DWT block, a) signal for head-checking defect undergoing DWT and determined dots with local minimums of function, b) coif5 wavelet along with decomposition level 16, c) number of samples with local minimums of function, d) the same as a for decomposition level 32 , e) for decomposition level 32 , f) as in c)

The values $d(k,n)$ are the discrete versions of $W(a,b)$ at $a = 2^k$, $b = 2^k n$. The DWT can continuously split the given original function f(t) into two parts: $f_0(t)$ corresponding to the coarser approximation of $f(t)$ and $g_0(t)$ corresponding to the high frequency detail function, defined as the difference between $f(t)$ and its approximated version. The approximated version $f_0(t)$ can be further split into two parts - the coarser approximation $f_1(t)$ and the detail part $g_1(t)$. This process can be continuously performed to any decomposition level.

The discrete wavelet transform can be obtained very efficiently by the application of so-called Mallat pyramid algorithm [11] and of two orthogonal (or biorthogonal) quadrature low-pass and high-pass filters. Different wavelet bases result in different forms of filters. According to the Mallat algorithm the low-pass and high-pass filtering are performed step by step producing the coarse approximation of the function and coefficients containing high frequency details, respectively.

The wavelet transform is able to separate the useful signal and the noise. The DWT accumulates the energy in a small number of DWT coefficients having large amplitudes and it spreads the energy of the noise over a large number of DWT coefficients having small amplitudes. Therefore the denoising process consists simply in removing DWT coefficients whose amplitudes are smaller than some assumed threshold value Th [11].

Based on it, authors attempt to check the usefulness of wavelet transform in the detection of head-checking defects. The experiment has been conducted in two stages.

In the first stage, high frequency components corresponding to the noise are removed, Fig. 3a and Fig. 3d.

In the second stage, the samples corresponding to head-checking flaws are searched in the denoised signal. This search is performed on the basis of local minimums of the denoised signal.

Authors tried to use many commonly available wavelets such as Harr, Daubechie, bioorthogonal, Cioflet [2], [11]. Attempts were performed for different decomposition levels affecting the threshold Th.

Fig. 3c and Fig. 3f present the effect of extraction of head-checking defect (marked as dots in Local Minimum field) for Coiflet5 wavelet and decomposition levels (Scale field) equal to 16 and 32.

The best filtering results have been obtained for Coiflet5 wavelet. Despite it, the results obtained are not completely satisfactory. Hence it seems that the design of an own wavelet matching the shape of signal corresponding to head-checking defect could improve the system performance.

4 Conclusion

The paper presents an attempt of fuzzy sets and wavelet transform application to the extraction of head-checking defects from signals generated by the method of scattering [8], [10]. On the basis of experiments performed, it is possible to say that the method based on fuzzy sets outperforms the method based on a wavelet transform. It results from following a priori assumptions. The voltage range of 0 – 2.5V for signals coming from head-checking defects and the voltage range of 3.5 - 10V for the noise have been assumed in the study. Using this assumption the method based on fuzzy sets is able to describe better the relationship relating the signal coming from head-checking defects with one of these sets.

The method based on a wavelet transform tries to find in the signal analyzed the components of assumed shape determined by the wavelet (for example Haar or Coif) and of assumed frequency dependent on the decomposition level (scale). Therefore acceptance of voltage level criterion at the extraction of head-checking defects significantly limits the usage of wavelet transform in the presented form. The wavelet transform could be useful if it is possible to determine the shape of actual signal coming from a head-checking defect. Then on the basis of this shape, it would be possible to design an own wavelet precisely describing the signal coming from head-checking defects. However it requires thorough and deep analysis of physical phenomenon occurring during the reflection of laser ray from the rail surface. It will be analyzed in the next stage of the work.

In further research the optimization of processing algorithms for better extraction of particularly dangerous flaws will be performed. Presented results from an example of rail part provide an introduction to further analysis of these methods applied to the diagnostics of railway rails.

These algorithms should be verified on the basis of investigations carried out on the railway lines, what in turn requires further work on the creation of professional mobile diagnostic system.

References

1. Cpałka, K.: Zagadnienia interpretowalności wiedzy i dokładności działania systemów rozmytych. EXIT, Warszawa (2009)
2. Daubechies, I.: Ten lectures on wavelets. SIAM Press, Philadelphia (1988)
3. Lesiak, P.: Mobilna diagnostyka szyn w torze kolejowym. Monograph No 116, Technical University of Radom, Radom (2008)
4. Lesiak, P.: Diagnostic technology of contact-stress flaws such as head checking in railway rails. Monograph, vol. 121, pp. 187–198. Technical University of Radom, Radom (2008)
5. Lesiak, P., Bojarczak, P., Migdal, M.: Inteligentne klasyfikatory wad kontaktowo – naprężeniowych w szynach kolejowych. Pomiary Automatyka Komputery w Gospodarce i Ochronie Środowiska (2), 13–17 (2009)
6. Lesiak, P., Migdal, M.: Cluster analysis of head checking flaws in railway rails subjected to ultrasound diagnostics. The Archives of Transport XXI(3-4), 51–66 (2009)
7. Lesiak, P., Szumiata, T.: Skaterometria laserowa wad head checking w szynach kolejowych. Pomiary Automatyka Komputery w Gospodarce i Ochronie Środowiska (2), 25–28 (2010)
8. Li, H., Torrance, K.E.: An experimental study of the correlation between surface roughness and light scattering for rough metallic surfaces. In: Duparré, B., Singh, Z.-H. (eds.) Advanced Characterization Techniques for Optics, Semiconductors, and Nanotechnologies II. Proceedings of SPIE, vol. 5878, SPIE, Bellingham (2005)
9. Łachwa, A.: Rozmyty świat zbiorów, liczb, relacji, faktów, reguł i decyzji. EXIT, Warszawa (2001)

10. Łukianowicz, C., Karpiński, T.: Scatterometry of Ground Surfaces. Measurement Science Review, Vol 3(sec. 3), 21–24 (2003)
11. Mallat, S.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. PAMI (1989)
12. Nguyen, H.T., et al.: A First Course in Fuzzy and Neural Control. Chapman&Hall CRC (2003)
13. Nowicki, R.: Rozmyte systemy decyzyjne w zadaniach z ograniczoną wiedzą. EXIT, Warszawa (2009)
14. Ogilvy, J.A.: Theory of Wave Scattering from Random Rough Surfaces. Adam Hilger, Bristol (1991)
15. Pedrycz, W.: Fuzzy Sets Engineering. CRC Press, Boca Raton (1999)
16. Ross, J.: Fuzzy Logic with Engineering applications. McGraw-Hill, Inc., New York (1997)
17. Stover, J.C.: Optical Scattering: Measurement and Analysis. McGraw-Hill, Inc., New York (1990)

Combined Simulated Annealing and Genetic Algorithm Approach to Bus Network Design

Li Liu[1], Piotr Olszewski[2], and Pong-Chai Goh[3]

[1] Shanghai Tat Hong Equipment Co., Beijing, PR China
lydia.liuli@gmail.com
[2] Warsaw University of Technology, Warsaw, Poland
p.olszewski@il.pw.edu.pl
[3] Nanyang Technological University, Singapore
cpcgoh@ntu.edu.sg

Abstract. A new method – combined simulated annealing (SA) and genetic algorithm (GA) approach is proposed to solve the problem of bus route design and frequency setting for a given road network with fixed bus stop locations and fixed travel demand. The method involves two steps: a set of candidate routes is generated first and then the best subset of these routes is selected by the combined SA and GA procedure. SA is the main process to search for a better solution to minimize the total system cost, comprising user and operator costs. GA is used as a sub-process to generate new solutions. Bus demand assignment on two alternative paths is performed at the solution evaluation stage. The method was implemented on four theoretical grid networks of different size and a benchmark network. Several GA operators (crossover and mutation) were utilized and tested for their effectiveness. The results show that the proposed method can efficiently converge to the optimal solution on a small network but computation time increases significantly with network size. The method can also be used for other transport operation management problems.

Keywords: Bus network design, optimization, genetic algorithm, simulated annealing.

1 Introduction

Improving the efficiency of public transport is an often-stated goal of transportation policy in big cities because only efficient public transport can successfully compete with private cars and thus help to ease the increasing traffic congestion. As buses are the backbone of public transport systems, optimization of bus routes would certainly contribute to improving the system efficiency. However, this problem is seldom tackled by transport planners. In most cases bus networks evolve incrementally – new services are being added as the city develops. In practice, bus route planning is often carried out by a combination of cognitive methods and trial-and-error.

In theory, optimization of bus routes for a city with a given road network, bus stop locations and passenger demand would involve deciding on the best number of routes and the best stop sequence for each route, with the objective of minimizing the sum of

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 335–346, 2010.

total passenger user costs and bus operating costs. It is basically a combinatorial problem and the number of possible solutions grows exponentially with the network size. Because of the problem complexity, literature on bus network optimization is relatively limited. Some early attempts to solve the problem analytically (e.g. [10]) involved greatly simplified networks and abstract models. The optimum values of key parameters could be determined by solving the objective function mathematically. However, real-life constraints on route length, route spacing, bus stop spacing and many-to-many character of demand cannot be reflected in these analytical models. This makes the analytical approach not practical for real life applications.

Another approach to solving the bus network design problem involves heuristic algorithms. Ceder and Wilson [3] considered both passenger and operator viewpoints and used a tree search algorithm to find all the feasible route sets. In practice, it is impossible for planners to examine all the solutions for a large network. It is helpful to use a screening algorithm, so as to find a limited number of better solutions before doing the final evaluation. Mandl [12] proposed an approach which started with a feasible set of routes and then searched for better solutions by trial and error. The improvement may be obtained by exchanging parts of routes, including a new node or excluding a node. Baaj and Mahmassani [2] proposed a similar hybrid route generation algorithm, starting with a simple skeleton network and extending it by adding new routes or inserting new nodes. The heuristic approaches have the advantage of using expert knowledge to reduce the search space.

With the increase of computer power, new approaches emerged for solving complex optimization problems with large search spaces. Genetic algorithm (GA) is one of the popular methods and has been applied to the bus route design problem [13]. Krishna Rao et al. [9] proposed a two-phase network design process using GA which was able to produce better results when compared with Mandl's study [12]. To validate the effectiveness of genetic algorithm, Chien et al. [4] applied the GA and exhaustive search to design a feeder bus route from an irregular service area with 160 zones. Identical optimal solutions were found by both methods but the GA was about 90 times faster. In order to extend the searching space, Lin et al. [11] suggested performing GA in parallel operations on three independent populations which were mixed and exchanged after each evolution. All of these studies proved that GA is efficient in searching for the optimum solution. However, it is known that GA may become stuck in a local optimum as it only accepts better solutions in its reproduction process. Yet accepting a worse solution is sometimes beneficial as it may lead to searching a wider space and finally reaching the global optimum.

Simulated annealing (SA) is also a promising algorithm with a strong search capability. However, its application in bus route design is very limited. Friesz et al. [6] applied SA to find an optimal network design when the flow pattern was constrained to be in equilibrium. In another study, a hybrid approach based on combing case-based reasoning and simulated annealing was proposed by Sadek [14]. Both of these studies showed that SA algorithm is efficient in solving optimization problems, although SA is not always capable of finding the global optimal solution. SA can sometimes accept worse solutions when certain requirements are met so as to extend its searching space beyond the local optimum.

It is natural to think that the combination of two powerful search algorithms may achieve better results, especially when one algorithm complements the other. Hence, a combined SA and GA approach is proposed in this paper. Studies on combined SA and GA are very rare. Zhao and Zeng [17] [18] used this method to search for an optimal solution for bus route network design problem. In their first study [17], the aim was to minimise the number of vehicle boardings that passengers have to make, while constraining the total route length and the number of routes. In their second study [18], Zhao and Zeng extended their method to optimise public transport network layout and headways with the objective of minimising total users' cost. They tested their method on Mandl's network and their solution was better than in previous studies. However, the objective function used by them is not quite practical as it ignores the bus system operating cost.

2 Problem Formulation

Bus route design involves not only planning of bus routes based on a road network, but also passenger demand assignment and frequency determination. Usually, the inputs to bus route design should comprise road network suitable for bus travel, bus stops locations and the stop-to-stop travel demand matrix. The aim of bus route design is to plan sufficient bus routes to cover all the bus stops and accommodate all the travel demand while not violating any constraints. Most common constraints are: fleet size, minimum and maximum service frequency etc. It is important to select a good objective function so that these constraints can be balanced. Based on a study by van Nes and Bovy [15], minimizing the total cost (the sum of operating cost and travelers' cost) is considered the desirable objective in bus network design with a fixed demand. This objective is also adopted in this study, i.e. the aim is to determine a set of bus routes that produces a minimum total cost while meeting all the requirements and constraints.

The bus route design problem can be formally presented in the following way. A road network $N = (Rd, B)$, comprises a set of roads $Rd = (1, 2 \ldots m)$ and a set of bus stops $B = (1, 2 \ldots n)$. A bus route is represented by a sequence of bus stop IDs. A set of candidate bus routes $R \{1, 2 \ldots k\}$ is generated for network N. Each bus route is given a unique number as its ID. A solution which is a subset of candidate routes, R_{SR} $\{1, 2 \ldots\}$, is represented by a sequence of bus route IDs. For example, a route network consisting of 5 bus routes: 4, 27, 34, 9, and 12, will be simply represented by a string with the length of 5 elements.

The objective function is to minimize the total bus system cost. This is expressed as follows (adopted from [13]):

$$Min : E = c_p \sum_{i=1}^{n} \sum_{j=1}^{n} T_{ij} \cdot t_{ij} + c_b \sum_{k \in R_{SR}} f_k L_k \ . \tag{1}$$

subject to the following constraints:

$f_k \geq f_{\min} \forall k \in R_{SR}$ (frequency feasibility)

$l_k \leq l_{\max} \forall k \in R_{SR}$ (load factor constraint)

where:

c_b unit bus operating cost per kilometer ($/bus-km)

c_p unit cost of passenger travel time ($/pass.-h)

E the value of objective function ($/h)
f_k frequency of k^{th} bus route (buses/h)
f_{min} minimum bus operating frequency (buses/h)
k bus route number
l_k load factor on k^{th} bus route = $Q_k^{max} / (f_k \cdot bus\ capacity)$
l_{max} maximum allowable load factor of bus service
L_k round trip distance for k^{th} bus route (km)
n number of bus stops in the network
T_{ij} travel demand from node i to j (pass./h)
t_{ij} total travel time (sum of in-vehicle time, waiting time and transfer time) between stops i and j (h)

The total travel time between bus stops i and j (t_{ij}) is the sum of in-vehicle travel time, waiting time and transfer time. The first term in Eq. 1 represents the total travel cost incurred by the traveler and the second term represents the total bus operating cost of all the routes in the solution route set. Hence, the sum of both terms represents the total system cost. This model will be solved using the proposed SA-GA approach.

3 Framework of the Proposed Method

The proposed hybrid SA-GA method is used to design bus routes. SA is used as the main algorithm to search for the optimal solution. In each SA iteration, GA is used to give a random disturbance to the current solution. Since every solution is actually a set of bus routes, it is necessary to generate enough candidate routes for each pair of bus stops in a road network before running SA-GA. This is done with the k-shortest path algorithm [8].

The main structure of SA-GA process is presented in Fig. 1. It is implemented in Visual C++ environment. The initial network N_0 is generated by a GA sub-function called *initialization* and its objective function value, E_0, is calculated. The maximum temperature (T) is set to 50. Parameter I is used to control the iterations of the GA process. The maximum number of iterations is set to 20. A new solution N_1 is obtained from the GA process and evaluated to obtain E_1. If the new solution is better than the old one (i.e. $E_1 < E_0$), it is accepted and N_1 replaces N_0. Even if the new solution is worse than N_0 (i.e. $E_1 - E_0 > 0$), it can still be accepted randomly, with the acceptance probability which is a function of the two objective function values and the current temperature (Fig 1).

GA is used to generate new potential solutions among which the one with the best fitness value is selected and subsequently used by the SA process for further selection [7]. In the case of minimization, the fitness value is usually assumed to be the reciprocal of the objective function value.

During the initialization phase, two tasks are accomplished. The first task is to generate a solution (a string of bus route IDs) N_0 which will be used as the initial solution. Since the size of the optimum solution is not known beforehand, variable string length coding is applied. The size of the solution *(s)* is selected at random within a predefined range and solution N_0 is generated by selecting s candidate routes at random.

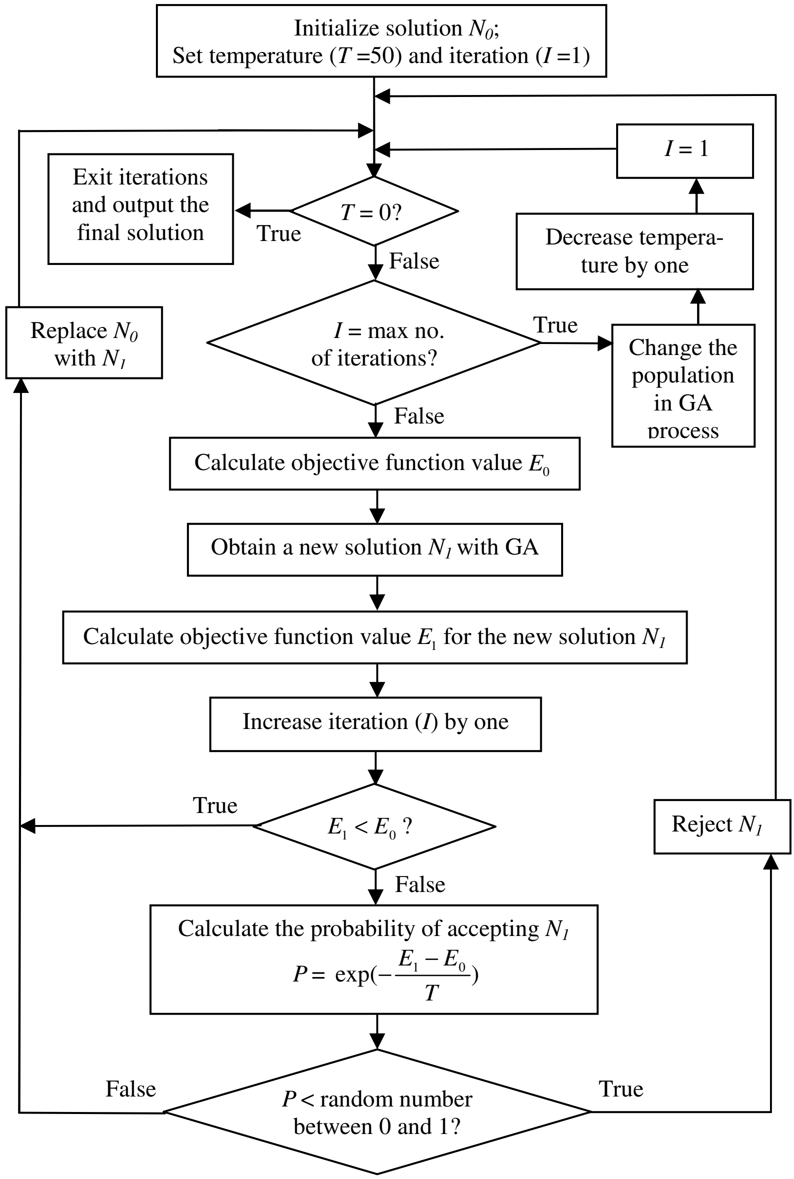


Fig. 1. Flowchart of the proposed SA-GA main process

The second task is to generate a population of solutions. Bigger population size gives better diversity and hence the chance of finding better solutions. However, bigger population increases the computation time. In this study, population size of 200 was chosen after the initial experiments showed that this did not increase the computation time dramatically. Using the same method as for initializing N_0, the other 200 solutions of the population are generated. Each individual solution is evaluated based on the objective function formula and its fitness value is determined for reproduction purpose in the GA process.

Reproduction is a process in which some strings with better fitness values are selected from the population for further GA operations. Before doing reproduction, N_0 is combined with the 199 strings to form a population with 200 strings. The popular roulette wheel selection method [7] is used in this study. In this way, strings with a higher fitness value have a higher probability of being selected.

Crossover is a process to generate new strings from a given pair of strings in the process of reproduction. For a set of strings, generally crossover is performed between the first and second strings, third and fourth, and so on. In addition, performing the crossover on a pair of strings is controlled by crossover probability P_c, which has a predefined value. To maximize the exploration space, P_c is set to 1 and crossover is performed on every pair of strings. Several crossover methods are available. Each has its characteristics and is potentially more efficient for certain types of problems. The common crossover approach applied in most GA applications is single-point crossover [7]. To test the effectiveness of different methods, three other kinds of crossover: two-point, uniform and convex are also used in this study.

Mutation is the occasional random alteration of the value of an element of a string. The mutation operator helps to introduce some potentially useful components (i.e. good bus stop sequences). As in crossover, this process is also controlled by a probability parameter P_m, (mutation probability). The frequency of mutation is usually very small in GA applications. One mutation per thousand elements is a typical value and produced best results in most applications. The value of $P_m = 0.001$ is also used in this study. Two types of mutations are used: uniform and dynamic. It should be noted that each type of mutation may result in duplicate elements in child strings. Since in the bus route network design problem no duplication is allowed, an additional procedure to remove duplicates is performed after a mutation. In this process, each duplicate is replaced by a route ID chosen randomly from candidate routes.

4 Application to Theoretical Networks

The SA-GA approach is applied to four theoretical networks to test its performance. The first network (N1) is shown in Fig. 2 and represents a two by three square grid with each road link 600 meters long. There are 14 bus stops (large dots) distributed on this network. Three of the bus stops are designated as terminals (crosses). These are the end points of all bus routes. Location of all bus stops is assumed fixed. The demand between all pairs of bus stops is also fixed. All bus stops, bus terminals and road intersections are considered nodes. Links between any pair of nodes are called bus links. It is assumed that all roads are two-directional. Three bigger theoretical networks (N2, N3 and N4) are generated in a similar way. Their characteristics are summarized in Table 1.

The task of candidate route generation should ideally be done with the k-shortest simple paths algorithm [8]. However, this algorithm has a high complexity and its computation time would be extremely large. Therefore, a modified Dijkstra algorithm [5] is applied to find the candidate routes between all pairs of bus terminals in order to save computation time. An amendment was made to this algorithm to exclude all paths with closed loops, as in reality a closed loop would normally not be acceptable as part of a bus route. It is not practical to ensure within an acceptable computation time limit that k candidate routes generated are the k shortest paths. However, the method guarantees that the shortest path is included as one of the candidate routes. The numbers of candidate routes generated for all four theoretical networks and the Mandl network are shown in Table 1. The process of route generation for all the networks is done within 30 seconds.

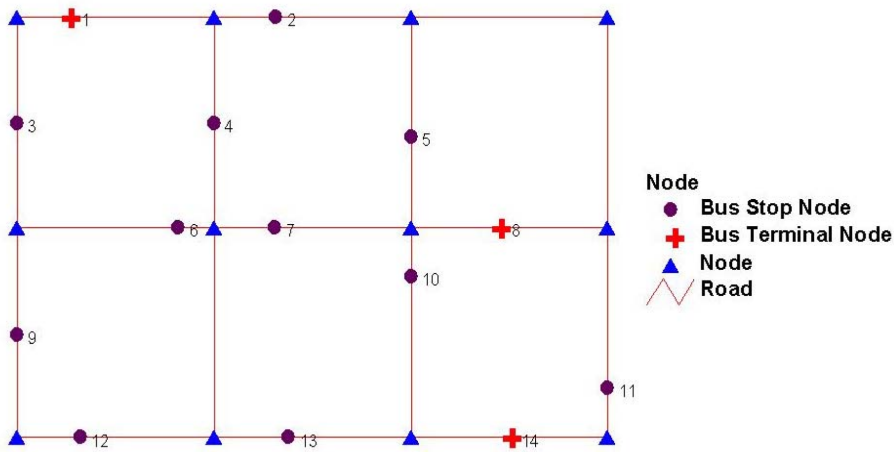


Fig. 2. Test network N1

Table 1. Characteristics of test networks

Network	No. of bus stops	No. of terminals	No. of nodes	No. of bus links	No. of candidate routes
N1	14	3	11	29	21
N2	20	4	14	40	171
N3	26	5	19	54	454
N4	32	6	23	67	720
Mandl	15	15	15	21	1183

To make the application more realistic, two-path assignment is used in this study. Two paths are considered between any two bus stops and used in travel demand assignment. Therefore, the objective function (Eq. 1) is modified as follows:

$$Min: E = c_p \sum_{i=1}^{n}\sum_{j=1}^{n} (T_{ij}^1 \cdot t_{ij}^1 + T_{ij}^2 \cdot t_{ij}^2 + T_{ij} \cdot t_{wij}) + c_b \sum_{k \in R_{SR}} f_k \cdot L_k \, z \,. \quad (2)$$

subject to the following constraints:

i. all bus stops are covered

ii. $f_k \geq f_{\min}$

where:

T_{ij}^1 travel demand from stop i to j assigned to the first path (pass./h)

T_{ij}^2 travel demand from stop i to j assigned to the second path (pass./h)

t_{ij}^1 total travel time for the first path between stops i and j (h)

t_{ij}^2 total travel time for the second path found between stops i and j (h)

t_{wij} average waiting time for travel form stop i to j (h)

The superscripts 1 and 2 represent the two paths used in the assignment – the first path is always the shortest and the second is an alternative path between stops i and j. The waiting time for passengers traveling from bus stop i to j (t_{wij}), is calculated as half of the headway determined considering buses forming the paths from i to j.

Usually, a frequency constraint is also imposed for the objective function. For a bus route with a frequency lower than the minimum frequency, the system will check whether the bus stop coverage constraint will be violated if this route was excluded from the solution. If the constraint is not violated, this route will be excluded and the first iteration of demand assignment is performed again. Otherwise this route will be kept and the frequency for this route is set to the predefined minimum frequency. However, a maximum frequency constraint was not used as this may exclude some very good routes which carry a lot of travelers. In real life, this problem can be solved manually by splitting one heavily loaded route into two similar routes.

Before performing the SA-GA, some parameter values need to be specified:

- Bus capacity: 80 persons/bus;
- Average bus speed: 18 km/h;
- Travelers' value of time c_p: \$8.00/hour [16];
- Bus operating cost c_b: \$5.24/bus-km (based on US public transport data [1]);
- Transfer penalty: 400 sec/transfer, assumed based on preliminary results.

Evaluating the solution is probably the most important and complicated part of the whole process. Since the solution is randomly generated and always changing, two sets of values need to be found before calculation of the objective function value and fitness based on Eq. 2. These are the in-vehicle travel times between all pairs of bus stops and frequencies for all bus routes in the solution. Frequency is obtained by dividing the maximum demand for each route by bus capacity. The demand for each route will keep changing while different routes are selected into the solution network. Therefore, it is necessary to do travel demand assignment every time a solution is formed or changed. The method adopted is to assign demand to the two selected paths. The rules for selecting the two paths are:

- the first path must be the shortest path between the pair of bus stops; and
- the in-vehicle travel time of the second path can not be more than 20% longer than the in-vehicle travel time of the shortest path.

Transit assignment is done in two iterations. In the first iteration, the demand between each pair of bus stops (T_{ij}) is assigned equally to the two selected routes and the frequencies for both routes are calculated. In the second iteration, the demand is re-assigned to the two routes according to their calculated frequencies.

Two options are considered when transfers have to be involved. One is direct transfer at the alighting bus stop. The other is an inter-connecting transfer by walking from the alighting bus stop to a different bus stop to transfer to another service. The inter-connecting limit is set to 300 meters of walking distance. The transfer penalty and walking time are added to the total path time.

The final frequency of each route can be calculated based on the resulting demand from the second transit assignment. The average waiting time for each trip can also be calculated. Now all the data needed for evaluation based on Eq. 2 are available.

The SA-GA approach as described above was first applied to network N1. Since four types of crossover and two types of mutation were used, in total eight different combinations were tested. The average computation time for one combination was 7 minutes for network N1. The overall best solution for network N1 is presented in Fig. 3. With this solution, only 38 stop-to-stop trips out of 182 possible combinations need one transfer in the shortest travel path. The others are all direct paths without transfer. The actual transfer rate is 7589/6434 = 1.18 transfers per trip. This means that 85% of trips were accomplished without any transfer.

Table 2. Summary of computation results of all networks

Network	Crossover used by the best result	Mutation used by the best result	Objective function value ($/h)	Solution size (routes)	Average time for SA-GA (min)
N1	Single/Two point/Uniform	Uniform/ Dynamic	10030	4	7
N2	Two-point	Dynamic	19986	6	37
N3	Single-point	Dynamic	36599	10	81
N4	Two-point	Dynamic	61698	19	133
Mandl (Min. time)	Two-point	Dynamic	26765	10	35
Mandl (Min. cost)	Two-point	Dynamic	24465	6	35

The best results for the other three bigger networks are presented in Table 2. With larger network size, the computation time has also increased significantly. The same procedure was applied to the benchmark network which was first presented by Mandl [12] and then used by many other researchers [2] [18]**.** The characteristics of Mandl's network were given in Table 1. To make the results comparable with the previous studies, two objective functions were used: minimising the total travel time and minimising the total cost.

The solution from SA-GA with minimisation of travel time as the objective, produced the result 37.8% better than Mandl's final solution and 22.6% better than Baaj and Mahmassani's best solution in terms of percentage of direct trips. In terms of total travel time, the result obtained is 17% better than Mandl's solution; 12% better than Baaj and Mahmassani's solution and 3% better than Zhao and Zeng's solution.

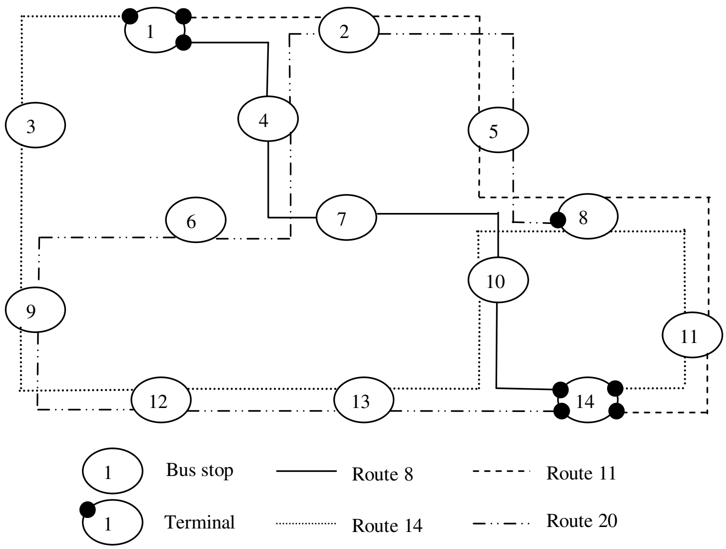


Fig. 3. Best solution for network N1

5 Discussion

The computation times of all the networks are indicated in Table 2. The computation time goes up very quickly with the increase of network size. This is not surprising because bus network design problem is a highly computationally-intensive problem (NP-hard). After examining the results of network N1, it seems very likely that an optimal or nearly optimal solution is found because the same solution is obtained in 5 combinations of mutation and crossover methods out of a total of 8. When comparing the features of the best results for all four networks, it can be seen that all the best solutions came from dynamic mutation and three best solutions came from the two-point crossover. This suggests that dynamic mutation and two-point crossover are more suitable than the other mutation and crossover techniques for solving the bus network design problems with GA.

Some of the parameters used for SA calculations like the starting temperature and number of iterations conducted at each temperature were selected in an attempt to balance the computation time and the size of search space. The selection of parameters of GA operations like population size, probability to conduct crossover (P_c) and probability to perform mutation (P_m) may affect the performance of SA-GA approach.

Since there is no scientific method for choosing the "optimum value" for these parameters, sensitivity tests were run to test the effectiveness of these parameter values. The best values obtained from sensitivity tests were used in the applications of SA-GA to both theoretical and the benchmark networks.

On the other hand, selection of some of the other parameters was tentative and may require further testing. For example, transit assignment was done in two iterations and only two paths were used. In the first iteration, demand was assigned equally to the two paths and in the second, assignment was based on bus frequency. The number of paths considered, the acceptable time range and the number of iterations were assumed in this way in order to control the computation time. Whether using three or more paths, setting larger acceptable time range or more assignment iterations would produce better results should be the subject of further investigation.

6 Conclusions

The bus route design problem is well known for its complexity and computational intensity as the number of possible combinations increases exponentially with network size. This paper has presented a combined SA-GA approach for bus route network design. It is implemented in two phases: a set of candidate routes is first developed and then an optimal subset of routes is selected. With the help of advanced algorithms: simulated annealing (SA) and genetic algorithm (GA), the design problem is solved in minutes. The SA-GA solution of Mandl's network is better than previous solutions and the in-vehicle travel time is quite near to its theoretical minimum. Moreover, a more realistic objective function and evaluation method are applied: demand is assigned to two paths for every pair of bus stops using an iterative procedure. This method has not been used in solving the bus route design problem before.

Results of the method application to four theoretical networks suggest that different genetic operators have different levels of efficiency. Since all the best solutions come from dynamic mutation and three best solutions come from two-point crossover, the indication is that dynamic mutation and two-point crossover are the most suitable for solving the bus network design problem with GA.

In terms of computation time, it is expected that the time will increase quickly and is approximately proportional to the cube of the size of the network. This result is not unexpected, as the bus route design problem has been proven to be one of the NP-hard problems which are nearly impossible to solve in polynomial time.

While the effectiveness of the proposed SA-GA approach has been demonstrated using theoretical networks, it is intended to further explore its effectiveness on a real bus network. In a case study, part of Singapore road network with around 400 links, 80 bus stops and 8 bus terminals will be used for bus route planning with SA-GA.

The proposed method can also be used for other transport operation management problems, for example: routing of demand-responsive transport, planning temporary bus routes for special events or cases when part of the network is closed due to maintenance or construction activities.

References

1. APTA: 2009 Public Transportation Fact Book. American Public Transit Association, http://www.apta.com/resources/statistics/Pages/transitstats.aspx
2. Baaj, M.H., Mahmassani, H.S.: Hybrid Route Generation Heuristic Algorithm for the Design of Transit Networks. Transportation Research Part C 3, 31–50 (1995)
3. Ceder, A., Wilson, N.H.M.: Bus Network Design. Transportation Research Part B 20, 331–344 (1986)
4. Chien, S., Yang, Z., Hou, E.: Genetic Algorithm Approach for Transit Route Planning and Design. J. of Transportation Engineering 127, 200–207 (2001)
5. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms, 2nd edn. The MIT Press, Cambridge (2001)
6. Friesz, T.L., et al.: A Simulated Annealing Approach to the Network Design Problem with Variational Inequality Constraints. Transportation Science 26, 18–26 (1992)
7. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Publishing Co., Reading (1989)
8. Hershberger, J., Maxel, M., Suri, S.: Finding the k Shortest Simple Paths: a New Algorithm and its Implementation. ALENEX, Baltimore (2003)
9. Krishna Rao, K.V., Muralidhar, S., Dhingra, S.L.: Public Transport Routing and Scheduling Using Genetic Algorithms. In: Proc. 3rd Int. Workshop on Transportation Planning and Implementation Methodologies for Developing Countries, IIT Bombay, pp. 91–102 (1998)
10. Kuah, G.K., Perl, J.: Optimization of Feeder Bus Routes and Bus-Stop Spacing. J. of Transportation Engineering 114, 341–354 (1988)
11. Lin, X.H., Kwok, Y.K., Lau, V.K.N.: A Genetic Algorithm Based Approach to Route Selection and Capacity Flow Assignment. Computer Communications 26, 950–960 (2003)
12. Mandl, C.E.: Evaluation and Optimisation of Urban Public Transport Networks. European J. of Operational Research 5, 396–404 (1980)
13. Pattnaik, S.B., Mohan, S., Tom, V.M.: Urban Bus Transit Route Network Design Using Genetic Algorithm. J. of Transportation Engineering 124, 368–375 (1998)
14. Sadek, A.W.: Hybrid Simulated Annealing and Case-Based Reasoning Approach for Computationally Intensive Transportation Problems. Transportation Research Record 1774, 18–24 (2001)
15. van Nes, R., Bovy, P.H.L.: The Importance of Objectives in Urban Transit Network Design. In: 79^{th} Annual Meeting of the Transportation Research Board, Washington D.C. (2000)
16. VTPI: Transit Evaluation. Online TMD Encyclopedia, Victoria Transport Policy Institute, http://www.vtpi.org/tdm/tdm62.htm
17. Zhao, F., Zeng, X.: Optimization of Transit Network Layout and Headway with a Combined Genetic Algorithm and Simulated Annealing Method. Engineering Optimization 38, 701–722 (2006)
18. Zhao, F., Zeng, X.: Simulated Annealing-Genetic Algorithm for Transit Network Optimization. J. of Computing in Civil Engineering 20, 57–68 (2006)

Application of Bayesian a Priori Distributions for Vehicles' Video Tracking Systems

Przemysław Mazurek and Krzysztof Okarma

Higher School of Technology and Economics in Szczecin
Faculty of Motor Transport,
Klonowica 14, 71-244 Szczecin, Poland
{mazurek,okarma}@wste.szczecin.pl

Abstract. Intelligent Transportation Systems (ITS) helps to improve the quality and quantity of many car traffic parameters. The use of the ITS is possible when the adequate measuring infrastructure is available. Video systems allow for its implementation with relatively low cost due to the possibility of simultaneous video recording of a few lanes of the road at a considerable distance from the camera. The process of tracking can be realized through different algorithms, the most attractive algorithms are Bayesian, because they use the a priori information derived from previous observations or known limitations. Use of this information is crucial for improving the quality of tracking especially for difficult observability conditions, which occur in the video systems under the influence of: smog, fog, rain, snow and poor lighting conditions.

Keywords: Intelligent Transportation Systems, video tracking.

1 Introduction

Intelligent Transportation Systems (ITS) [11] consist of the following subsystems: the measurement, the control algorithms, executive subsystem, and the infrastructure for monitoring and data transmission. One of the most important parts of the ITS is the measurement subsystem. For the estimation of vehicles' parameters in this part of the ITS system different kinds of road sensors are used. The most important of them are the inductive loops and the VIPs (Video Image Processors) [11]. Presently, it is possible to control the road traffic and perform some monitoring tasks using such sensors. Obtained measurement results are further processed in order to influence on road users using the traffic lights and road signs with variable contents. the VIPs use analog or digital cameras and process the acquired images using some digital image processing and 2D pattern recognition algorithms.

There are many advantages of camera based ITS but the most important is the possibility of mounting without without interfering with the road surface. Furthermore, it is also possible to perform the measurements not only for a single vehicle but for a number of them at once. Some parameters of the vehicles can be obtained only by the VIP systems and they are further used for the classification

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 347–355, 2010.

of vehicles by shape and color. There are also some limitations of VIPs and the most important is the placement of the camera. The measurement range of such systems can be limited to the near or distant vehicles. Form the practical point of view the most interesting placement of the camera is top mounting where the camera's look point is located vertically downwards. Such mounting is possible only when a pylon is used, which is located near to the road, or on the wall of high buildings (that is possible only in the urban area, mainly the centers of large cities). The measurement range depends additionally on atmospheric conditions and unfortunately this important parameter is reduced at night and in the presence of fog, snow or smog.

2 Video Based Techniques for Vehicles' Tracking

The vehicles' tracking task is a kind of multitarget tracking [2], [3]. It is relatively simple for top mounted cameras, since all visible vehicles are well separated on acquired images. Unfortunately, such configuration is almost impossible outside the urban centers, so some occlusions may occur for most other mounting types and therefore some more advanced software techniques are necessary. Alternatively, the use of multiple cameras allow tracking also for low mounted cameras, but some multisensor tracking techniques are necessary for such cases due to required data fusion. Nevertheless, for the simplification of the analysis a single top view camera, or similarly mounted, is assumed in this paper.

One of the most often used techniques in tracking systems are the Kalman filter [10], [6] due to its optimality for linear sensor observation, Gaussian measurement errors, and Gaussian motion models [12]. The Kalman filter is not optimal for the motion typical for the road vehicles and is limited for the linear motion trajectories. There are some extensions of the Kalman filter like EKF [12] but such single filter is insufficient for such applications. If several possible trajectories or some nonlinear ones are expected, multiple filters are necessary for tracking. In such situation the tracking can be processed by IMM (Interacting Multiple Model) algorithms, which support the process of switching the trajectories within an assumed set for particular time moment and current position. This switching process can be controlled by computing of all filters' results (prediction) and their comparison with measurements. In such algorithm the minimal distance error can be used as the criterion of switching. For example a road with multiple curves could be modeled using a set of linear and curved trajectories. The process of switching could be also controlled by the position of the vehicle, which should be known for a specific road. Such technique is not used for air traffic but could be used for land and marine transportation.

For more sophisticated cases, the Bayes filter can be used, because of supported non–linear sensor observations, Non-Gaussian measurement errors and non–Gaussian motion models. The Bayes filter has also a very important advantage over the Kalman filtering approach, since the Kalman filter needs contact measurements (e.g. the position of the vehicle if this vehicle is visible). It means that the additional processing of measurement (detection) is necessary and in the simplest case threshold based techniques are applied. The Bayes filter supports contact and sensor measurements. The sensor usually acquires raw or simple preprocessed signals with non–binary information. Therefore, the signal values from sensors can be used more efficiently if the binarization is not performed. In the case of weak signal of tracked object (e.g. a vehicle tracked in poor atmospheric conditions) the difference between the vehicle's signal and background could be insufficient to distinguish them. The Bayes filter supports also such cases and tracking in the low SNR (Signal–to–Noise Ratio) conditions is possible. Such systems are generally known as a Track–Before–Detect approach [4], [5].

The Bayesian approach differs from others by the application of the objective a priori knowledge. Such knowledge could be incorporated to the tracking algorithm in two ways: as direct a priori values or as some boundary conditions (e.g. limitations of possible trajectories). A proper selection of a priori knowledge [1] gives the possibility of much better utilization of the measurements even if they are noised or if there are not enough of them.

The Bayesian approach supports recurrent data processing so the obtained results are used in next step. Recurrent data processing is common for non–trivial tracking algorithms (also like the Kalman filter). The prediction of the vehicle's position is necessary in a noised environment and multiple vehicles traffic. The predicted position is compared with the obtained measurements and used by the motion estimator's update [4]. The Bayesian approach could process road traffic measurements using conceptually simple steps, but in some specific cases (like in the TBD systems) it needs a lot of computation power.

3 The Bayesian Approach for Tracking

The Bayesian approach is based on the Bayes Theorem and the conditional probability density for X given the observation $Y = y$ is

$$p(x|y) = \frac{L(y|x)p(x)}{\int L(y|x)p(x)dx} \tag{1}$$

where p is the probability function for X.

It can be rewritten as:

$$p(x|y) = C^{-1}L(y|x)p(x) \tag{2}$$

where C can be omitted, since the maximum value is searched.

The sequential character of processing can be illustrated using two consecutive processing steps. The first processing step:

$$p(x|y_1) \propto L_1(y_1|x)p(x) \tag{3}$$

and the second one:

$$p(x|y_1, y_2) \propto L_2(y_2|x)L_1(y_1|x)p(x) \propto L_2(y_2|x)p(x|y_1) \tag{4}$$

so the last element $p(x|y_1)$ is the new a priori value in the second iteration.

for the higher number of states x:

$$p(x_i|y_1, y_2) \propto L_2(y_2|x_i)p(x_i|y_1) \tag{5}$$

and the possible states can be considered as the hypotheses. In the data analysis for tracking purposes the hypotheses can he the high–level trajectories, including several time steps, or the low–level ones, describing the characteristics of the measurement sensor. It is also possible to process some already detected locations of vehicles or processing the data from the sensor in order to detect the vehicle's location and determine its trajectory.

A specified hypothesis describes that a given location (pixel) is related to a vehicle or its fragment if a sensor uses more pixels as a representation of the vehicle, moving in a specified direction with a given speed. In such case y represents the pixel's value.

An interesting fact is that the a priori knowledge in the sequential meaning is not only the value of $p(x)$ in the first step but also the likelihood from the previous one. In the next steps, the influence of the new data y_i increases in comparison to the initial a priori distribution $p(x)$.

A description of the vehicle on a road can be performed with a finite but very large number of trajectories covering even several hundred time steps. This type of approach is quite troublesome, since it requires large computational power. Assuming that a current hypothesis depends only on the previous one there is a possibility of using the Markov matrix for the description of the process, which decreases the amount of computations.

This transformation makes the current a priori knowledge the motion update integral, providing information related to the possible transitions. The example transformation of a priori land avoidance function (for the marine example described above) to the Markov matrix can be found in [12]. The basic recursion for single target tracking is expressed as [12]:

Initial Distribution:

$$p(t_0, s_0) = q_0(s_0) \tag{6}$$

for $s_0 \in S$ where S is the state space.

For $k \geq 1$ and $s_k \in S$

Motion Update:

$$p^-(t_k, s_k) = \int q_k(s_k|s_{k-1})p(t_{k-1}, s_{k-1})ds_{k-1} \tag{7}$$

Information Update:

$$p(t_k, s_k) = C^{-1}L_k(y_k|s_k)p^-(t_k, s_k) \tag{8}$$

EndFor

where s_k – particular state space cell (e.g. position and corresponding velocity and direction), k – time step.

If the the objects are well separated in the state space, presented approach can be also used for multiple vehicles' tracking. Such technique can be applied mainly for the cameras mounted directly over a road.

4 Motion Limitations by a Priori Knowledge

Application of a priori knowledge should be used for the detailed analysis of the situation on the road. In the book [12] an interesting example is discussed in order to demonstrate the advantages of Bayesian approach filter in comparison to the results obtained by Kalman filter. A submarine or ship is used as the example, tracked wrongly using Kalman filter, since the obtained result would mean that the ship flows on the land, which is obviously impossible. The application of Bayesian filter with a priori knowledge preserves such situations. This example is very eloquent and sometimes cited also in publications about the road traffic.

Regardless of the fact that in case of water transport the example is true, it cannot be directly transferred to the other transport systems. This would be a serious error because the vehicles move not only on the road (even the movement on some pedestrian walkways is physically possible). Such problems are especially important and well visible in case of urban traffic. One of the most important urban ITS tasks is not only the motion tracking but also the detection of some potentially dangerous situations. Such incidents are particularly important for the ITS, since they may not be limited to signaling some potential dangerous behaviors of the drivers, but can also detect abnormal situations, not necessarily properly interpreted by the system. For example some taxi drivers relatively often perform risky maneuvers on the road. The reasons may be related to their desire to earn money or just to avoid waiting.

In a typical approach it is assumed that the vehicles move on the road and the probability of the presence e.g. on a specified lane can be determined. Observing the roads in the real situations it can be easily noticed that some vehicles appear on the bus passes, tram trackways, reserved areas, bike roads etc. Some of the vehicles are privileged but in some situations such behaviors are related to the violation of traffic rules.

Such situations are not very often, so for the static algorithms it is equivalent to the low possibility of the vehicle's presence in a specified area. In many systems such low probability it is ignored but for the ITS such information can be very important, also in terms of traffic safety. The detection of such incidents allows the real-time reaction, automatic informing the specified departments and the path optimization using e.g. the variable content road signs.

The above analysis proves the necessity of the extension of the observation in the ITS also outside the road surface and for this purpose many various approaches can be applied e.g. rough sets, fuzzy logic etc.

5 Comparison of Bayesian Approach for Conventional and Extended Motion Vectors

For the comparison of conventional approach (vehicles are present only on the road) and extended one (vehicles are present also outside the road) two tracking examples will be shown as the illustration of the importance of a priori knowledge. The avoidance function can be incorporated as time varying one, but in

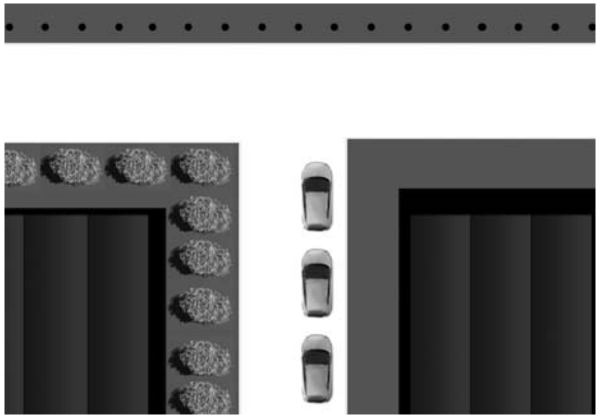


Fig. 1. An example scenario used in our simulations

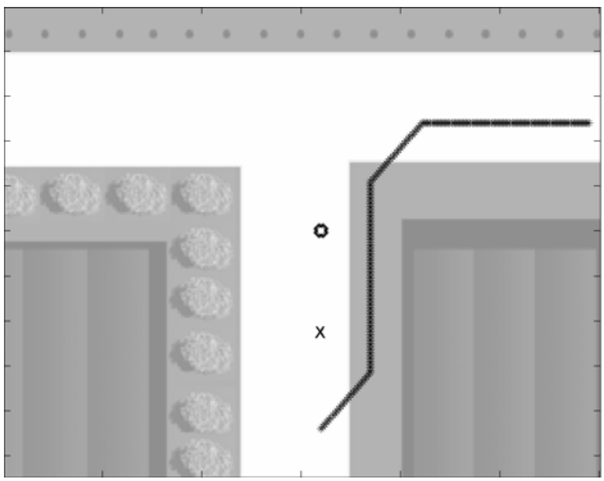


Fig. 2. The synthetic trajectory of the moving vehicle

this particular example the fixed one is assumed. There is no avoidance function among the vehicles because a physical distance separates them on the road.

The extension of the a priori limitations is not free, since only a fragment of the road and its surroundings is processed with this assumption. Some reliable limitations are necessary in order to avoid some extreme cases when a vehicle would move e.g. through a tree or a wall.

The detection of a vehicle is performed using another algorithm leading to the high value of $L_k(y_k, s_k)$. An example scenario of the crossroads is illustrated in Fig. 1, where three vehicles are stopped and the third one turns right by the pedestrian walkway while the two other stand at the crossroads. The synthetic trajectory of the third vehicle is shown in Fig. 2.

Fig. 3. The illustration of motion vector's weights and the limitations for the standard algorithm - sequence: LEFT, ZERO, UP(+12,+36), UP(+3), RIGHT

Fig. 4. The illustration of motion vector's weights and the limitations for the extended algorithm - sequence: LEFT, ZERO, UP(+12,+36), UP(+3), RIGHT

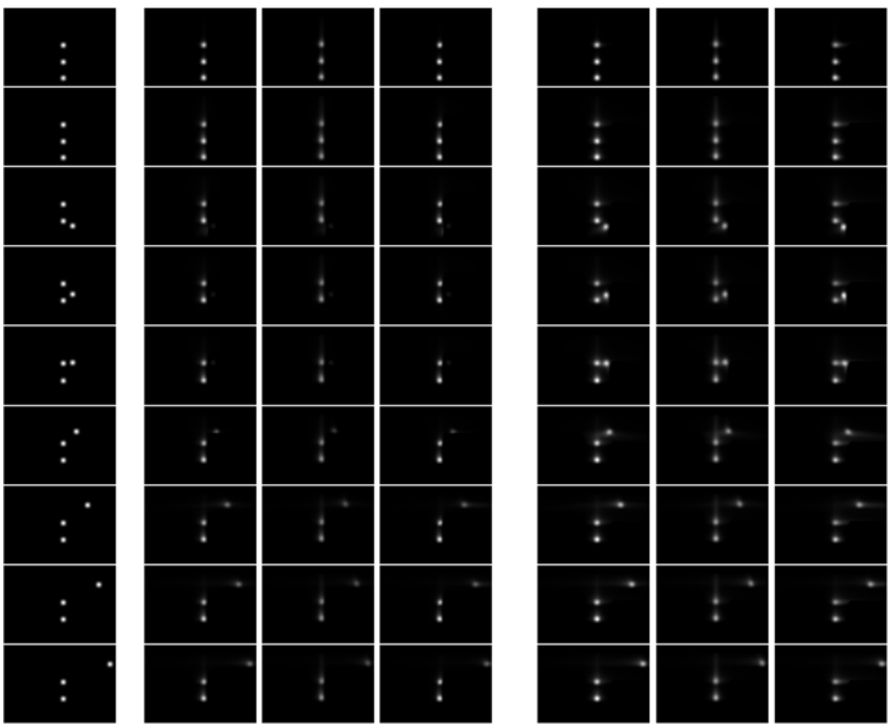


Fig. 5. The illustration of the state space for the conventional and extended algorithm - sequence: ZERO, UP(+3), RIGHT(+3)

Th emotion is described using the motion vectors: at the place (one vector), up (+3, +12, +36), left (+3, +12) and right (+3, +12). Several transitions between different states have been defined with high probability of transitions (typically the remaining in the current state corresponds to the values from 0.6 to 0.9). Furthermore, the scope of possible transitions has been described using several maps. The zero value (black color on the map) means the prevention of the transition and white color means the full speed range (vector lengths for a given motion vector). Grey values indicate the intermediate states. Such approach allows a flexible description of motion and for some larger objects (more than one pixel) the number of necessary motion vectors decreases together with the computational time. The extended algorithm covers also the area usually not used by the vehicles. Assuming additionally that relatively low speed is possible for the case of the UP motion, various UP vectors have been used.

The visualization of the state space is difficult because of its 4-dimensional character [7], [8], [9] and changes after each step, so only some chosen states for the most representative motion vectors are illustrated in Fig. 5. The left side illustrated the conventional approach, while the right one - the extended one. In the first case the object is lost for the moment (rows 2–5) because of the presence on the restricted area (reserved for pedestrians). Since the Markov matrix introduces the correlation among the data, the object (especially a large one) may appear in each of the motion vectors.

6 Conclusions

The use of the simple analysis based only on the motion for the detection of the object's presence at specified location may be inefficient. Using a priori knowledge based e.g. on Markov transitions some atypical situations can be handled, increasing the overall performance of the ITS. Analysis of such untypical transitions is possible due to the combination of data acquired as a result of standard analysis (typical trajectories and transitions) with extended Bayesian one based on a priori knowledge. Such analysis may be helpful for the analysis of the traffic safety, especially in some dangerous places, possibly leading e.g. to some changes of the traffic organization.

Acknowledgments. This work is supported by the Polish Ministry of Science and Higher Education (Grant No. N509 399136 „Estimation of the vehicles' motion trajectories using the Bayesian analysis and digital image processing algorithms“).

References

1. Aczel, A.D.: Complete Business Statistic. Richard D. Irwin (1993)
2. Bar–Shalom, Y.: Multitarget–Multisensor Tracking: Applications and Advances, vol. II. Artech House, Norwood (1992)

3. Blackman, S., Poupoli, R.: Modern Tracking Systems. Artech House, Norwoood (1999)
4. Blackman, S.: Multiple–Target Tracking with Radar Applications. Artech House, Norwood (1986)
5. Boers, Y., Ehlers, F., Koch, W., Luginbuhl, T., Stone, L.D., Streit, R.L. (eds.): Track Before Detect Algorithm. EURASIP Journal on Advances in Signal Processing, Hindawi (2008)
6. Brookner, E.: Tracking and Kalman Filtering Made Easy. Willey–Interscience, Hoboken (1998)
7. Coué, C., Pradalier, C., Laugier, C.: Bayesian Programming Multi–Target Tracking: an Automotive Application. In: Int. Conf. on Field and Service Robotics, Lake Yamanaka, Japan, July 14–16 (2003)
8. Coué, C., Fraichard, T., Bessiére, P., Mazer, E.: Using Bayesian Programming for Multi–Sensor Multi–Target Tracking in Automotive Applications. In: Int. Conf. on Robotics and Automation, Taipai, Taiwan, May 12–17 (2003)
9. Gindele, T., Brechtel, S., Schröder, J., Dillmann, R.: Bayesian Occupancy Grid Filter for Dynamic Environments Using Prior Map Knowledge. In: Proceedings of the IEEE Intelligent Vehicles Symposium (2009)
10. Kalman, R.E.: A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME–Journal of Basic Engineering 82 Series D, 35–46 (1960)
11. Klein, L.A.: Sensor Technologies and Data Requirements for ITS. Artech House ITS library, Norwood (2001)
12. Stone, L.D., Barlow, C.A., Corwin, T.L.: Bayesian Multiple Target Tracking. Artech House, Norwood (1999)

Evaluation of Early Jet Lag Symptoms by Passengers Crossing 7 Time Zones

Rafał Bobiński and Anna Michalik

Faculty of Health Sciences, University of Bielsko-Biala,
Konopnickiej 6, 43-300 Bielsko-Biala, Poland
rbobinski@ath.bielsko.pl

Abstract. Numerous studies have shown that travelling across three or more time zones may lead to jet lag, which is a consequence of circadian misalignment that occurs after crossing time zones too rapidly for the circadian system to keep pace. The main aim of this paper was to define the early symptoms of jet lag in travelers crossing seven time zones in the western and eastern direction. In order to verify the hypothesis there were 175 subjects, who could evaluate subjective severity of jet lag using a visual analogue scale. The current study showed that the significant predictors of early jet lag are tiredness, moodiness and gastrointestinal disturbance. There is a correlation between subjective effects of jet lag and physical well-being, tiredness, bowel movements, concentration and irritability.

Keywords: Jet lag, predictors of early jet lag, circadian rhythms.

1 Introduction

The jet lag syndrome is a disorder occurring as a result of endogenous maladjustment of sleep and wakefulness rhythm to the requirements of external environment, resulting from the changes of time zone [1-6]. It was classified to sleeping disorders connected with circadian rhythms. The diagnostic criteria of jet lag include insomnia and excessive sleepiness during the day connected with transcontinental flights. What is more, there is an activity impairment during the day, general fatigue and somatic symptoms (gastrointestinal disorders) maintaining 1 to 2 days after the flight [7-10]. Sleeping disorder is not a result of other disorders of the organism (neurological, mental, pharmacological) and it is only a result of sudden change of a few time zones [7]. The time needed to resynchronize circadian rhythms is dependent on the journey destination and what is connected with it, on the number of crossed time zones. It is shown in the research that flights involving crossing 3 or more time zones can contribute to the occurrence of the syndrome of sudden time zone change [5].

Inner desynchronization of circadian rhythms is responsible for general malaise, tiredness, night sleep disturbance, poor mental efficiency, emotional disorders as well as gastrointestinal problems, which usually occur in the first week after long-distance flights. Sleeping disorder may include problems with falling asleep and sustaining sleep what takes place after flights from east to west, and waking up too early in the case of flights from west to east [11]. The occurrence of symptoms is dependent on

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 356–363, 2010.

individual predispositions and therefore, they can be felt immediately after reaching the airport, although their enforcement in the second and next nights after the flight is more typical for jet lag.

Among early symptoms one can enumerate growing tiredness, weaker ability to concentrate, higher level of nervous excitability as well as digestive disorders. These symptoms can be significant in evaluating the strength of jet lag syndrome in next days.

2 The Aim of the Paper

Along with the increasing interest in travelling there are more people choosing long-haul flights, crossing several time zones. Because of that, research have been conducted which have been aimed at exploring the jet lag syndrome, however they usually concentrate mostly on symptoms occurring in the second and next days after reaching the destination. Nevertheless, knowledge about its early form is insufficient. The aim of this paper was to define early symptoms of jet lag, occurring in the travelers crossing 7 time zones in eastern and western direction, to evaluate the influence of chosen factors and the conditions of flight on the jet lag effect and to evaluate the travelers' knowledge about jet lag.

3 Material and Methods

The testing tool was a self-generated questionnaire. It consisted of two parts: the first – which enabled general characteristic of the surveyed group (age, sex, nationality of the respondent and information about coffee and alcohol intake before and during the flight) and the second – including questions about the main issue of the research – subjective evaluation of jet lag by the means of a visual analogue scale (VAS). The level of fatigue sensation, the work of digestive system as well as the severity of experiencing jet lag were evaluated on a scale from 0 meaning lack of disorders to 10 showing a great severity of its sensation. Mental and physical well-being, fatigue, hunger, satiety, irritation and concentration were evaluated on a scale ranging from -5 (worsening, lowering, weakening) to +5 (improving, intensifying).

The questionnaire was prepared in Polish and English version, taking into account the multinationality of the respondents.

The research was conducted among 175 people, who were divided into three groups. The first of them consisted of passengers of the flight Chicago – Cracow (36 people). The questionnaires were given right after the arrival in the luggage-reclaim area. The second and the third group of the respondents filled in the questionnaire: during the last 30 minutes of the flight Warsaw – Chicago (57 people) and Chicago – Warsaw (83 people).

Standard statistical methods were used to investigate and describe the results.

4 Results

The majority of respondents were women (57.71%). When the sex of the respondents was analyzed according to the variant of the questionnaire it was proved that the differences are not significantly differential.

When it comes to age, the biggest group of the respondents consisted of passengers aged 30-39 (25.71%) and then respondents aged 40-49 (23.4%).

The age according to the questionnaire variants is presented in Fig 1.

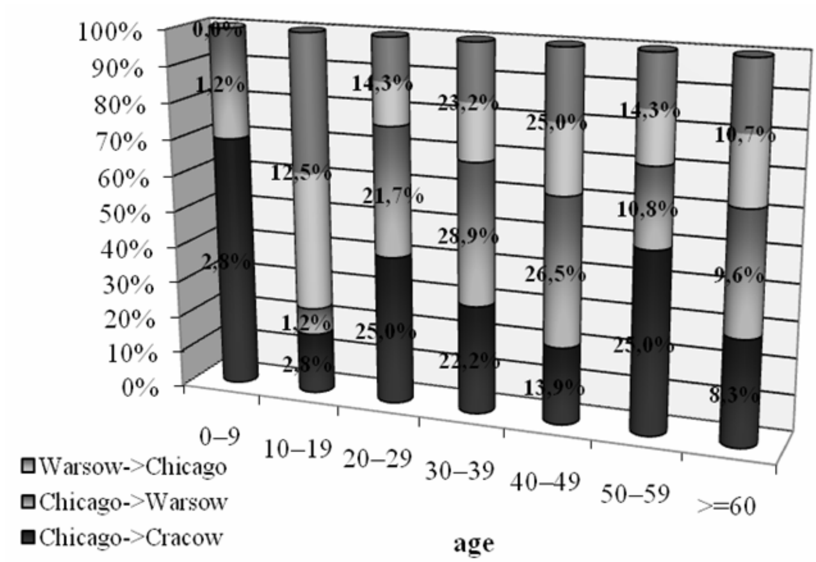


Fig. 1. The age of the respondents according to questionnaire variants. Taking into account the flight variant it was shown, that factions are not importantly varied.

Evaluating the awareness of the existence of jet lag among the respondents it was proved, that 55.43% have heard about the jet lag syndrome, what is shown in Fig. 2.

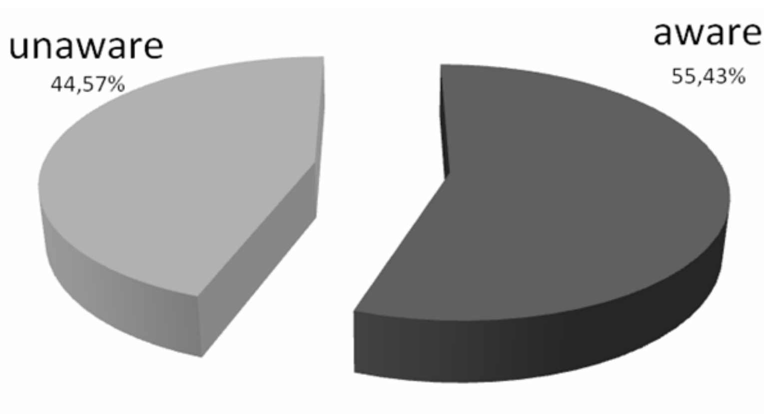


Fig. 2. Awareness of the jet lag syndrome among respondents

It was also noticed that there is no significant dependence between possessing knowledge about jet lag and feeling its early symptoms. In people, who were aware of the existence of jet lag the average of sensing its early form was 2.777. In the group unaware of jet lag this sensation amounted to 2.915.

Based on the evidence, the analysis of reliability and position of the variables was conducted, determining Alfa-Cronbach's, and we tried to construct an aggregate variable describing an early jet lag effect. The reliability reached Alfa level 0.7 for each particular symptom (Table 1).

Table 1. Descriptive statistics for each particular early jet lag symptom

Early symptoms of jet lag	N	Average	Median	Minimum	Maksimum
Physical well-being	175	-0,069	0	-5	5
Mental well-being	175	0,194	0	-4	5
Fatigue	175	2,091	1	0	9
Hunger	175	0,337	0	-5	5
Satiety	175	0,160	0	-5	4
Bowel activity	175	0,960	0	0	10
Irritation	175	0,463	0	-5	5
Concentration	175	-0,269	0	-5	5

It was shown (correlation according to Spearman rank) that significant in describing the effect of early jet lag are tiredness, evaluation of the bowel activity and irritation, what is shown in Fig 3.

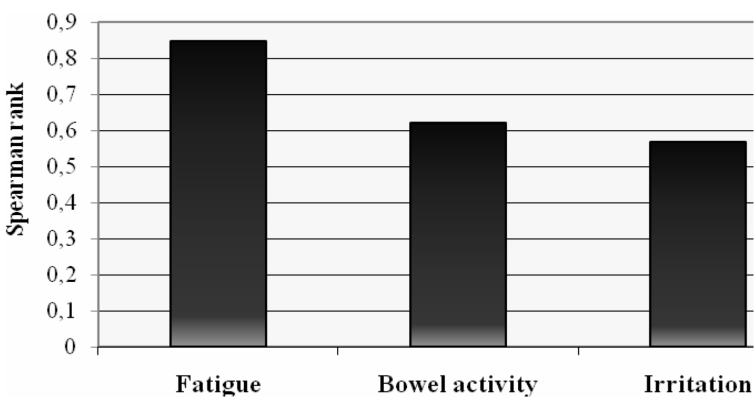


Fig. 3. Components of early jet lag

Using the analysis of multiple regression of jet lag variable verifying the influence of such variables as variant of the questionnaire, sex, age group, awareness of jet lag on intensifying early effect, it was shown that only variant A of the questionnaire (flight Chicago – Cracow) has significant adverse influence on early jet lag symptoms (Fig. 4).

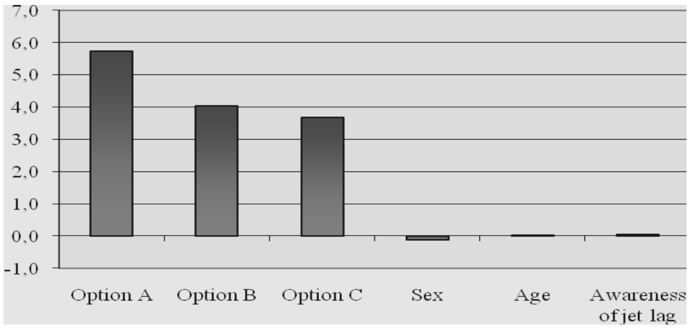


Fig. 4. Variables influencing jet lag

In the confirmatory analysis with the use of Kruskal-Wallis Anova scale it was shown that the effect of early jet lag is significantly differential when it comes to the variant of conducted questionnaire A (Chicago – Cracow), B (Chicago – Warsaw) and C (Warsaw – Chicago). In the post hoc analysis it was stated that variant A is significantly varied from the others. The average sensation of early jet lag in this variant amounts to 4.306 (Fig. 5).

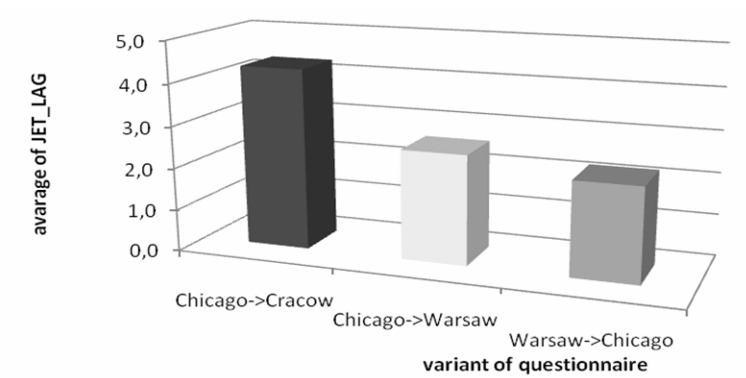


Fig. 5. Jet lag and the variants of questionnaire

The analysis of ranked correlation of jet lag and components of early jet lag showed that the average of its sensation amounts to 2.094. The order correlation according to Spearman rank proved that such components as physical well-being, fatigue, evaluation of digestive system, irritation and concentration are significant to self-evaluation of early jet lag. Those correlations are significant when reaching probability not smaller than 0.5 (Fig. 6).

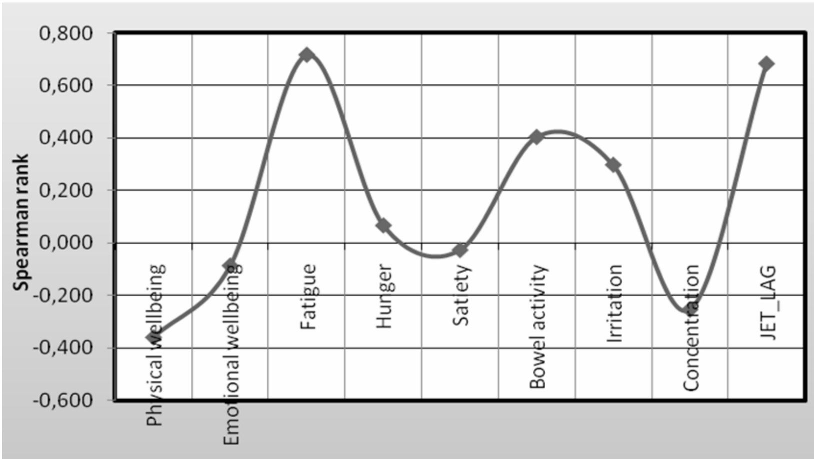


Fig. 6. Correlation between jet lag and its symptoms

Analyzing the questionnaire with the use of aggregate variable and Mann-Whitney test, in order to define correlations between early jet lag and sex, it showed that intensification of jet lag in men is insignificantly higher (2.881, women 2.779).

The research also revealed that intensification of early jet lag was insignificantly higher in people using alcohol. People drinking alcohol before and during the flight on average estimated the level of early jet lag sensation as 3.117, whereas in people not drinking alcohol it was lower (2.756).

When people were drinking more coffee than usually, it was observed that experiencing jet lag syndrome right after landing was insignificantly lower, namely 2.532. In people who did not drink coffee or drank less than usually, the average jet lag sensation was 2.935.

5 Discussion

Along with the increasing interest in transcontinental journeys the risk of the occurrence of the jet lag syndrome is also rising. In order to minimize the symptoms connected with the syndrome of sudden time zone change it is indispensable to inform the travelers about pathophysiology of jet lag. However, own research showed that the awareness of travelers about jet lag syndrome is unsatisfactory – 44% of travelers are not aware of the existence of jet lag.

The rising interest in long-haul flights also contributed to the need of conducting research aimed at studying the jet lag phenomenon, however it usually focuses on symptoms occurring in the second and next nights after reaching the destination [12], [13]. Insufficient, however, is the knowledge about its early form, therefore it has become the topic of the following paper. Own research showed that the place and conditions, in which the questionnaire was conducted have a significant influence on results received and they vary when it comes to different variants of the questionnaire. It was shown that only type A (flight Chicago – Cracow), has a significant influence on early jet lag sensation. This variant was the only one conducted straight after the arrival, in the luggage-reclaim zone, where the travelers faced external factors for the first time, which could trigger resynchronization of the biological clock. What is more, awaiting the luggage exerted a stressful influence on the respondents, what could be observed in their reluctance and irritation, and even in aggressive attitude towards pollsters.

Variant C of the questionnaire (flight Chicago – Warsaw) and variant B (flight Warsaw – Chicago) turned out to be statistically irrelevant. The questionnaires were handed to the passengers on the board during the last 30 minutes of flight (no contact with external environment) and showed, that differences concerning the level of sensation of early jet lag symptoms in people travelling on the same rout but in opposite directions are statistically irrelevant.

Summing up, it can be observed that first jet lag symptoms can occur as soon as in the first hour after landing, what was also pointed out by Waterhouse in his publications. Research also showed that such factors as fatigue and irritation are significant when it comes to early jet lag evaluation. Tiredness is the variable which is felt the strongest. Research done by Marks showed, that subjective tiredness can occur even in the fifth hour of flight [14]. Own research showed, however, that the level of fatigue during the flight is an early jet lag symptom of little significance. This corresponds with Waterhouse' reports, who categorized tiredness to significant factors in later jet lag sensation [5], [11], [12], and [13].

Digestive problems are the next factor correlating with jet lag, what is in opposition to other publications, which consider the relation between jet lag and gastrointestinal problems as a loosely linked and unimportant jet lag indicator [11]. From the discussion with passengers at the airport in Cracow – Balice, it can be concluded that, significant and high evaluation of this variable can be attributed to the fact of the occurrence of constipation, which can result from long immobilization and diarrheas connected with the stress caused by the journey. Moreover, passengers could suffer from nausea and vomiting caused by air sickness.

In literature, the correlation between level of irritation and jet lag is varied from irrelevant to significant [5], [11], [12], and [13]. Own research showed significant influence of early jet lag on irritation. It could be connected with the place and conditions in which the questionnaire was conducted. The only relevant variant of questionnaire – A (flight Chicago – Cracow) was handed to the respondents while awaiting their luggage, what could have had an influence on its results.

The correlation according to Spearman rank showed, in addition, a close correlation with self-evaluation of the above mentioned variables and additionally with physical well-being and concentration. It corresponds with other research, which showed that the ability to concentrate is the second factor (right after tiredness) strongly connected with the jet lag syndrome [5], [11], [12], and [13].

It is also worth drawing attention to other factors, such as age, alcohol and coffee intake on early jet lag. In many studies, the age is listed as a variable reinforcing the jet lag syndrome. It concerns mostly the elderly [5]. Own research showed lack of adverse influence of age on the perception of early jet lag symptoms. Coffee and alcohol consumption also has insignificant influence on the sensation of early jet lag symptoms. Attention should be drawn however, to the fact that the effects of its influence could appear at a later stage. Therefore, it is advisable to limit the alcohol intake which could contribute to dehydration of the organism as well as it could intensify fatigue and jet lag [5].

6 Conclusion

Obtained results made it possible to come up with the following conclusion:

1. Over half of the respondents (55.43%) have heard of jet lag, however access to information about this topic is still insufficient.
2. Significant components of early jet lag are tiredness, irritation and gastrointestinal disorders.
3. Self-evaluation of the effect of early jet lag correlates with physical well-being, tiredness, bowel activity, ability to concentrate and irritation.
4. The only relevant variant of the questionnaire is variant A (flight Chicago – Cracow), conducted straight after landing in the luggage reclaim zone.
5. Intensified jet lag in men, people drinking alcohol and in people, who did not drink too much coffee, is insignificantly higher.

References

1. Brown, G., et al.: Melatonin and its relevance to jet lag. Travel Med. Infect Dis. 7, 69–81 (2009)
2. Coste, O., Lagarde, D.: Clinical management of jet lag: What can be proposed when performance is critical? Travel Med. Infect Dis. 7, 82–87 (2009)
3. Lu, B.S., Zee, P.C.: Circadian rhythm sleep disorders. Chest 130, 1915–1923 (2006)
4. Zisapel, N.: Circadian Rhythm Sleep Disorders, Pathophysiology and Potential Approaches to Management. CNS Drugs 15(4), 311–328 (2001)
5. Waterhouse, J., Reilly, T., Atkinson, G., Edwards, B.: Jet lag: trends and coping strategies. Lancet 369, 1117–1129 (2007)
6. Reilly, T., Waterhouse, J., Edwards, B.: Some chronobiological and physiological problems associated with long-distance journeys. Travel Med. Infect Dis. 7, 88–101 (2009)
7. Sack, R.L.: The pathophysiology of jet lag. Travel Med. Infect Dis. 7, 101–110 (2009)
8. Barion, A., Zee, P.C.: A clinical approach to circadian rhythm sleep disorders Sleep Med. 8, 566–577 (2007)
9. Waterhouse, J.: Jet-lag and shift work (1) circadian rhythms. J. R. Soc. Med. 92, 398–401 (1999)
10. Auger, R.R., Morgenthaler, T.I.: Jet lag and other sleep disorders relevant to the traveler. Travel. Med. Infect Dis. 7, 60–68 (2009)
11. Waterhouse, J., et al.: The relationship between assessment of jet lag and some of its symptoms. Chronobiol Int. 20, 1061–1073 (2003)
12. Waterhouse, J., Neville, A., Finnegan, J., Williams, P., Edwards, B., Kao, S., Reilly, T.: Further assessments of the relationship between jet lag and some of its symptoms. Chronobiol Int. 22(1), 121–136 (2005)
13. Waterhouse, J., et al.: Do subjective symptoms predict our perception of jet lag? Ergonomics 43, 1514–1527 (2000)
14. Marks, E., et al.: Multi-shift work and work under conditions of jet lag. Questionnaire studies. Med. Pr. 33(1-3), 125–127 (1982)

Selected Issues of a Coordinated Adaptive Road Traffic System Application within the Silesian Conurbation

Renata Żochowska, Ireneusz Celiński, Aleksander Sobota, and Leszek Czapkowski

Silesian University of Technology, Department of Traffic Engineering, Krasinskiego 8, 40019 Katowice, Poland
{Renata.Zochowska,Ireneusz.Celinski,Aleksander.Sobota, Leszek.Czapkowski}@polsl.pl

Abstract. The article deals with the selected issues of deployment of a coordinated adaptive road traffic control system within the Silesian conurbation. One needs to identify all peculiar problems allocated with this region. This will include the range, structural and administration issues. All of them are associated with the distributed traffic management realized by different local authorities. Since deployment of such complex system relies mainly on the ITS solutions, all important technical issues applying to the conurbation are being presented. The authors have specifically highlighted a problem of collecting information on travel destinations. The efficient systems in use base on the dynamic methods and need the Origin-Destination matrices to be updated due to a current situation on roads. All the conventional methods of estimating such matrices do not apply and one needs to consider the assumptions on and limitations of the input data and complexity of computing.

Keywords: ITS, ATMS, ATCS, SCATS, SCOOT, traffic model, OD matrix.

1 Introduction

Coordinated adaptive road traffic control systems in Poland have been deployed since the 90s of the 20th century and most of them are based on the Sydney Coordinated Adaptive Traffic System (SCATS) [25]. First implementations of such road traffic control systems in Poland were to be found in Cracow, Poznan, Rzeszow and Warsaw [26]. The SCATS system is one of the two most popular solutions used in a field of the coordinated adaptive road traffic control in the world. It is widely used in Asia and Oceania. Sydney, for that instance, has 3700 intersections being controlled by the system, and Hong-Kong has put the SCATS into use on 1300 intersections. In Europe it is very popular in the Republic of Ireland. There are 500 intersections in Dublin under operation of the system. Currently the SCATS system operates approximately on 23000 of intersections in 100 countries worldwide.

Another solution in the coordinated adaptive road traffic control widely used in many places worldwide is the SCOOT system. It is in service on a large scale in the United Kingdom as well as in such cities as Beijing and Bangkok.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 364–373, 2010.

Both systems use different methods employed in the control algorithms. In addition, the basic traffic characteristics formed in each system, such as values of the time losses and a number of the stopping, do not differ more than 11% [9], [14], [18].

The coordinated adaptive road traffic control system in Poland has been applied to:

- Wrocław – 180 intersections;
- Warsaw – 160 intersections;
- Poznań – 112 intersections;
- Cracow – 80 intersections;
- Lódź – 61 intersections;
- Olsztyn – 33 intersections;
- Gdynia – 33 intersections;
- Rzeszów – 5 intersections.

The Silesian Conurbation does not have any coordinated adaptive road traffic control system of any kind deployed as of now. The main metropolitan area is inhabited by approximately 2.6 million people. Taking also a demographical potential of the Rybnik agglomeration area with its 500 thousand inhabitants into consideration, a prospective system of the coordinated adaptive road traffic control might be applied to a region inhabited by over 3 million people. Therefore, deploying such system in such area might be compared in terms of operation to the already functioning systems in the biggest European agglomerations.

2 Coordinated Adaptive Road Traffic Control System in the Silesian Conurbation

2.1 Range Issues

When talking about deployment of the coordinated adaptive road traffic control system to the Silesian Conurbation one needs to take the range of such project into consideration. Hence, the range issues need to be pointed out. This is strictly related with the main road traffic corridors within the conurbation (Fig. 1) [11].

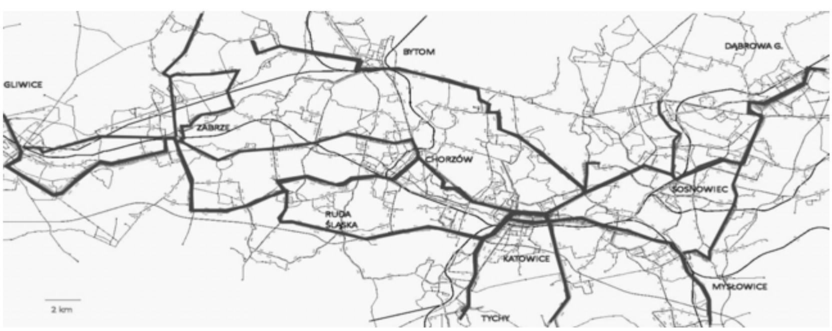


Fig. 1. Main road traffic corridors within the Silesian conurbation [11]

A road network of the Silesian Conurbation serves mostly the eastbound and westbound traffic as a result of geographical layout of the region. It also serves heavy northbound and southbound flows. This of course is the main reason of the increase of traffic in the network and vehicles accumulating on the streets, roads and junctions. As a result of such process the bottlenecks seem to naturally appear throughout the entire network [27].

A complex system of the coordinated adaptive road traffic control shall cover the westbound and eastbound traffic corridors between Gliwice and Dąbrowa Górnicza on a distance of 60km. It shall also cover a distance of 45km of the northbound and southbound corridors between Tarnowskie Góry and Tychy. That makes just about 2700 square kilometers of the surface area to cover. When taking also the Rybnik agglomeration into consideration, then the distance of the northbound and southbound traffic corridors will increase up to 90km.

Deploying the system of the coordinated adaptive road traffic control in such area might be compared in terms of operation to such European agglomerations as London or Paris. Thus, some range issues exist in terms of deployment as in any road network system the characteristics of the traffic delays are nonlinear [27]. The delays of the *i* traffic flow aggregate of the time losses, also at the intersections within the system. These time losses describe all vehicles within a range of the coordinated adaptive road traffic control system as well as those out of range. There are no any technical and economical means to fully deploy the Intelligent Transportation Systems to such road network. Suggestions of a solution to the issue have been made in [12]. And the problem of range could be described by a mathematical formula as following [27]:

$$w_i(q_1, q_2, \ldots, q_n) = \sum_{j=1}^{l_i} w_i^{(j)}(q_1, q_2, \ldots, q_n), \quad l_i \geq 1 \ , \tag{1}$$

where:

$w_i(q_1, q_2, \ldots, q_n)$ – delay of the *i*-traffic flow;

$w_i^{(j)}(q_1, q_2, \ldots, q_n)$ – delay of the *i*-traffic flow at the *j*-intersection;

n – a number of traffic flows;

l_i – a number of intersections passed through by the *i*-traffic flow;

q_i – intensity of the *i*-traffic flow.

2.2 Structure Issues

It is the local authorities such as City Streets and Bridges Management or Roads Management that manage the road networks within the urban areas in Poland [19], [20], [21] and [22]. One of the problems appearing in the Silesian Conurbation is structuring of the road network. This is done by the financial departments within local authorities responsible for maintaining the roads. It has of course a huge influence on molding the traffic flows as well as the traffic characteristics in any road network. It is also a huge problem being usually openly ignored, and good examples of such are the intersections without traffic lights. Such intersections, by all technical and economical means, are quite often excluded from the analysis appearing in the systems of the coordinated adaptive road traffic control. However they can be used in those systems as it has been shown in [8].

The percentage of intersections without traffic lights for the urban road networks in Poland is big and therefore, it is becoming a very important issue to include them as a vital part of the road network when deploying the coordinated adaptive road traffic control systems. The percentage of intersections without traffic lights for the selected towns and cities of the Silesian Conurbation is as following [12]:

- Katowice – 52.66;
- Mysłowice – 29.79;
- Chorzów – 28.92;
- Bytom – 25.2;
- Siemianowice Śląskie – 21.95;
- Sosnowiec – 18.75;
- Tychy – 7.51;

The above inventory shows main diversification of contribution of the particular technical solutions when it comes to the road traffic management within the conurbation. Some of the cities and towns, just by their location in the metropolitan area, should have a statistically crucial number of the intersections with the traffic lights in their own road networks. However, such forms of the traffic management are not put into practice at all.

Tychy, comparing to other towns and cities of the conurbation, has a much bigger number of the roundabouts than any other town or city of the entire region. They serve the most important traffic corridors in town and for that reason there is no need to employ there the intersections with the traffic lights. So by deploying the coordinated adaptive road traffic control system within the conurbation and including Tychy in such, it might be needed to go with the solutions out of the ordinary when it comes to the coordinated adaptive road traffic control [12].

2.3 Transport Policy Issues

Deploying the system of the coordinated adaptive road traffic control to the Silesian Conurbation needs to be done after an initial yet very important stage takes place. This applies to gaining the knowledge of and collecting all vital data about the issues regarding such implementation and the solutions to these within a centralized system of the coordinated adaptive road traffic control. This of course could determine not only which system of the coordinated adaptive road traffic control to choose from (SCATS, SCOOTS, TACTICS, ACTRA etc) but also which of the technical solutions of the ITS subsystems to opt for (Scheduling&Dispatch, GIS, Electronic Fare Collection, AVL, ATI, IVA, APC, SP, VMS etc). As of now there is no centralized transport policy in terms of a widely understood road network development in the Silesian Conurbation.

The traffic surveys, as well as the full traffic surveys (KBR), run independently by each local authority, are very often out of date (Katowice – last survey took place in 1998) or performed inaccurate (Tychy, 2008, incomplete traffic model). Such situation is mostly caused by the economical factors. Some of the local authorities cannot afford to run such surveys, even though they are obligated by the law to do that on a regular basis. There are of course some traffic models made for the conurbation but they had basically been created with regard to some other technical implementations and solutions, such as a tramway network [11]. Hence, there is no traffic model as of now that would cover all the towns and cities as well as the most important traffic corridors within the Silesian Conurbation. One of the most problem-causing issues regarding status quo is lack of any administrative body—although the Upper Silesian Metropolitan Union could be seen as such—managing the entire conurbation and therefore lack of integrated transport policy for the region.

There is an entity managing the public transportation in the Silesian Conurbation and that is the Municipal Transport Union of the Upper Silesian Industrial District (KZK GOP). Nonetheless there is also a number of small independent entities operating in the region and running public transportation. That causes some issues with the coordination of the transportation offer for passengers. Some issues with the possibilities of establishing a common policy on fares and ticketing between different city transit authorities appear as well. An undergoing process of metropolization of the Silesian Conurbation should solve such issues for years to come. Quite recently a regional railway passenger transport has been cast under control of the local authorities, which could be seen as an example of a growing regional integration in terms of the common transport policy. It is expected that developing a metropolis within the Silesian Conurbation will cause a full integration of all entities running the public transportation services and will reach a stage where one fully integrated entity would perform such services for the entire region. This would definitely make the implementation of the coordinated adaptive road traffic control system much easier and would apply not only to the road traffic.

2.4 Technical Issues

When implementing the ITS solutions to the coordinated adaptive road traffic control, one needs to take the Adaptive Traffic Management products into consideration. However, prioritizing public transportation on this area may be extremely difficult as the road network is very dense. There is not enough of the urban space throughout the conurbation for the bus lanes to be segregated or new ones to be built. This is due to the narrow streets of the 19th century in the city centers as well as the narrow downtown layouts of the period in general. Lack of beltways in the majority of towns and cities as well as a small number of the alterative traffic corridors for specific destinations does not make it any easier.

A prior trend regarding a contribution of the public transportation services for the entire metropolitan area is reversing [11]. All travels by the modes of public transportation are dropping down. People are having a tendency to travel using private transport as being more convenient to the commuter and available on demand [11]. It might however overwhelm the capacity of the transportation network in the region and would lead to the increase in congestion. One also needs to expect that without regard to which of the ITS solutions have been chosen, when it comes to the public transportation, it will be necessary to implement an information system, such as the *real-time bus information at the stop*. It will allow all the means of public transportation to be fully used in terms of their capacity. This applies of course not only to trams and buses but also to trains. Especially at present as the railway transport modality has much bigger capacity reserves than its road counterpart. Only 7% of all travels in the conurbation in 2009 were realized by rail [11]. The information system should operate based on one of the ITS solutions, such as *Scheduling&Dispatch*, a vehicle positioning system for public transportation.

2.5 Stages of Implementation

Implementation of the coordinated adaptive road traffic control system based on the ITS solutions should be done in two stages. First the ITS technologies need to be applied from a point of the identified processes regarding congestion. Such implementation will increase the efficiency of the road network, however without any change to the characteristics of these processes (*suboptimization*). Hence these technologies are applied with regard to those identified problems, particular and limited by space and time.

The second step is being brought to solve the issues causing the congestion in traffic by the traffic modeling and analysis. It can be done by integrating the specific ITS solutions as well as introducing new solutions of the ITS systems, such as increasing the area covered by the centralized traffic control, which on the other hand helps to find additional, better suboptimal solutions [12]. This is important because of the bottlenecks are relocating throughout the dense road network of the conurbation. These relocations are caused by the changes in a distribution of the traffic flows in the road network by the spatial migration of the traffic generators and absorbents in a longer time period [11]. In addition, the coordinated adaptive road traffic control system in operation can also cause some relocation of the bottlenecks, even by the Signal Priority subsystem as well as the periodic breakdowns associated with the operation of the system. Such disruptions can be also caused by a long time of implementation of the coordinated adaptive road traffic control system and its calibration.

3 Dynamic Origin-Destination Matrices – Basic Input Data for the Coordinated Adaptive Road Traffic Control System When Congestion in Traffic

The main goal of implementing the road traffic management system is to reduce the congestion in traffic and to maximize the capacity of the transportation network. In order to get effective results one should—based on available data—constantly observe a changing structure of the traffic. It is a consequence of decisions taken by passengers and mostly concerning the mode of transportation and the route. The knowledge about the influence of the traffic management parameters on the infrastructure usage is helping to increase the quality of the prognostic models that are used in all dynamic processes.

The area of the Silesian Conurbation is characterized by a large disproportion of the weight of traffic. There are two types of the traffic congestion to be distinguished in general [16]: structural and incidental. Due to the structural congestion it is possible to determine the origin-destination travels as a function of time as well as corresponding alternative routes. The incidental congestion comes into existence as a result of the breakdown, road works and road accidents. In the traffic modeling it can be used to determine the spatial sensitivity, expressed in a form of a resistance function. The function essentially depends on the experiences of the road users, knowledge of the network and information received from the traffic management systems. This information can be handed over to the passengers in two ways: as the roadside information or as the onboard information.

Estimation of the Origin-Destination (OD) matrices is a key element of the adaptive layer of the hierarchical road traffic management systems [1]. With an effective emergency management for the road accidents, breakdowns or road works the knowledge of destination of each traffic participant is necessary. Advanced computing algorithms create the OD matrices based on the current traffic volume on a road and velocity of each particular vehicle. The matrices are being constantly updated in defined time intervals and soon after the update takes place, the information is being projected on the network. A real situation is then compared to a computer simulation that is being systematically generated. With every substantial deviation the short-term prognostic models are being appropriately adjusted. Only then it is possible to route the optimal diversions in the urban road networks.

The centers of gravity for each region (centroids) represent the points of accumulation of the potential of the origin and destination regions. In all practical solutions these centers are usually shifted by—so called—the connections to the closest nodes of a technical network such as road or railway network. Therefore, it is possible to determine the OD matrix between two regions by establishing the traffic volume between two nodes of the technical network.

There are two completely different methods of determining the OD matrix [4]: static and dynamic. The static approach takes a balance within the network in a longer time horizon for granted and it is usually applied for the larger networks. Each cell of a two-dimensional OD matrix represents the traffic volume, expressed by a number of travels between two regions.

The dynamic approach, as an alternative, deals with the variability of the traffic flows. A problem of determining the dynamic OD matrix is completely different to a problem of the Dynamic Traffic Assignment [8], [17], [23] and [13]. Basically a solution is to find the OD matrix that after distributing it onto the transportation network will lead to the traffic structure similar to the real one. A graph describing the transport network structure in time and space has been expanded by introducing a support matrix that, for each time period, determines a number of intervals needed to pass a network section by a vehicle at this section [5]. Using such interpretation it is possible to reduce the dynamic problem to the static one that is set into a particular time-space continuum.

Practice brings also the hybrid solutions. They use data in a form of the time series to determine dynamically the values of the traffic flows. They also use more aggregated data to determine the OD matrix structure on a network level. In case of rare or non-crowded networks, determining the OD matrix and its distribution onto the transportation network is going independently. In that case correlations between the OD matrix and the traffic flows can be described by a set of equations resulting—among others—from a theory of the maximal entropy and minimal information [28], a generalized method of the smallest squares [3], [6], Bayesian inference [15] or a method of the greatest credibility.

For the complex networks, such as for the Silesian Conurbation one, the short time horizons are mandatory. All the methods known by far determine the distribution onto the network according to the rule of constant proportions and therefore they are suitable for the non-crowded networks [2], [7]. For the networks characterized by the congestion in traffic it is not possible to determine the OD matrix and therefore to conduct the distribution onto the network independently. This leads to the two-level programming, in which both issues are the sub-programs solved in a sequence [10], [29]. A method suggested by Sherali, Sivanandan and Hobeika [23] is more beneficial as it uses the principles of stability when measuring the traffic flows at the network sections. This can be then further aggregated until the OD matrix is received. This method has been further developed including non-linear programming and stochastic processes [4].

In a real life situation a set of the network sections, for which the values of the traffic volume in the following time moments have been taken, is often much smaller than a set of all sections. This leads to the indeterminacy of the operational research task and so various methods to estimate the missing elements are being applied. Additional data is to be associated with the technical and traffic characteristics of the nodes and sections. Moreover, the initial OD matrix depends mostly on a chosen computing algorithm. Quite often it is a value determined by previous demand matrices (gotten empirically or analytically) or estimated due to the traffic volumes having been observed.

Choosing an objective function is not that simple. It usually shows how far real values of the traffic volume taken at the network sections are from a copybook solution. This “distance” is being minimized in a computing process. The objective function is usually being specified for all sectors but it also can be established only for a critical section with its highest deviations from the empirical values. The initial value of the objective function is often established based on the initially accepted OD matrix.

For each section and node one assumes a certain value of the resistance function that depends on variability of the traffic volume in time. There are few methods to establish the variability: to start with assumption on fixed factors and to finish with complex calculation methods.

Algorithms used to find the shortest paths within the network are the most complex in terms of computing and such complexity even increases with the enlargement of the network. In the dynamic methods the computation time is vital and despite the technological development in terms of carrying out complex calculations, one tends to use the algorithms of the lowest computing complexity. For each destination within the OD matrix a path with a minimal momentary value of the resistance function is being selected. Next a degree of use by vehicles in a given destination is being determined. In dense networks, at some destinations in particular, a few or even several paths are being selected so the appropriate volume of the OD matrix could be distributed.

Real time control requires the prediction of the value of the traffic parameters for the next control period. Therefore, a proper predicting of the spatial distribution for the following time horizons is crucial. It is in order to distribute the predicting matrix in such way that is most accurate with the real situation. The historical matrices and observing the traffic conditions evolving in time are very useful in this process.

An important issue when it comes to the dense transportation networks, is safety that should particularly in the traffic management systems be taken into consideration. Surveillance of the transportation network in order to monitor the road accidents as well as analysis of their occurrence should generate a strategy of the information and traffic control in case of any event causing the growth of congestion.

Further research should concentrate on the dynamically and constantly updated replies from the system. Due to the current values of the traffic volume it shall choose the best solution for travels in all destinations in terms of minimization of the time and costs. Also monitoring the influence of the control elements on behavior of passengers is significant as it is to evaluate the size of the traffic flow transferred on the alternative paths.

4 Conclusions

The increase of the traffic jamming in Polish cities and towns intensifies the congestion in traffic and therefore it directly corresponds to a standard of life of all their residents. For that reason there is a constant need to improve the quality of operation for both, the public and the private transport. These issues are becoming really very important, especially for all metropolitan areas and therefore for the Silesian Conurbation as well as it is covered with a dense road network and a huge inconsistence in terms of the traffic volumes. One of the methods to improve such situation is to deploy and implement a coordinated adaptive road traffic control system. And so these issues are these days one of the most essential challenges for the traffic engineering and traffic engineers.

References

1. Adamski, A.: Inteligentne systemy transportowe: sterowanie, nadzór i zarządzanie. Uczelniane Wydawnictwa Naukowo-Dydaktyczne Akademii Górniczo-Hutniczej, Kraków (2003)
2. Ashok, K., Ben-Akiva, M.E.: Dynamic origin-destination matrix estimation and prediction for real-time traffic managements systems. In: Proceedings of the 12-th International Symposium on the Theory of Traffic Flow and Transportation, pp. 465–484. Elsevier, Amsterdam (1993)
3. Bell, M.G.H.: The estimation of origin-destination matrices by constrained generalized least squares. Transportation Research B 25, 13–22 (1991)
4. Bell, M.G.H., Shield, C.M., Busch, F., Kruse, G.: A stochastic user equilibrium path flow estimator. Transportation Research C 5(3/4), 197–210 (1997)
5. Bierlaire, M.: The total demand scale: A New measure of quality for static and dynamic origin-destination trip tables. Transportation Research 36 B, 837–850 (2002)
6. Cascetta, E.: Estimation of trip matrices from traffic counts and survey data. Transportation Research B 18, 289–299 (1984)
7. Cascetta, E., Inaudi, D., Marquis, G.: Dynamic estimators of origin-destination matrices using traffic counts. Transportation Science 27, 363–373 (1993)
8. Cremer, M., Keller, H.: A New class of dynamic methods for the identification of origin-destination flows. Transportation Research 21 B, 117–132 (1987)

9. Davies, P.: Assessment of advanced technologies for relieving urban traffic congestion, National Cooperative Highway Research Program 340. Transportation Research Board, NRC Washington D.C. (1991)
10. Florian, M., Chen, Y.: A bilevel programming approach to estimating O-D matrix by traffic counts. Report CRT-750, Centre de recherché sur les transports, Montreal (1991)
11. Karoń, G., Macioszek, E., Sobota, A.: Selected problems of transport Network model ling of Upper-Silesian Agglomeration (In Poland), Vilnius Technika VGTU, Vilnius (2009)
12. Krawiec, S., Celiński, I.: Sterowanie obszarowe - przykłady rozwiązań w aspekcie modelowania ruchu drogowego, Międzynarodowa Konferencja Naukowa-Transport XXI wieku, Białowieża (2010)
13. Li, B., Moor, B.: Recursive estimation based on the equality-constrained optimization for intersection origin-destination matrices. Transportation Research 33 B, 203–214 (1999)
14. Liu, D. Cheu, R.L.: Simulation Evaluation of Dynamic TRANSYT and SCATS-Based Signal Control Logic under Time Varying Traffic Demand. Applications of Advanced Technologies in Transportation Engineering (2004)
15. Maher, M.J.: Inferences on trip matrices from observation on link volumes: a Bayesian statistical approach. Transportation Research B 17, 435–447 (1993)
16. Muller, T.H.J., Miska, M.P., Van Zuylen, H.J.: Monitoring traffic under congestion. Base for dynamic assignment in online prediction models. TRB 2005 Annual Meeting, Washington D.C (2005)
17. Nihan, N.L., Davis, G.A.: Recursive estimation of origin-destination matrices from input/output counts. Transportation Research, 21 B, 149–164 (1987)
18. Ngoc Nguyen, V.: Evaluation of SCATSIM-RTA Adaptive Traffic Network Simulation Model. Transportation Research Record: Journal of the Transportation Research Board (2007)
19. Ustawa o samorządzie gminnym z dnia 8 marca 1990 r., Jednolity tekst Dz. U z 2001 r. nr 142 poz. 1591 z późn. zm. art. 18 ust.2 pkt 9 lit. h
20. Ustawa o drogach publicznych z dnia 21 marca 1985 r., Jednolity tekst Dz. U z 2004 r. nr 204 poz. 2086 z późn. zm. art. 21 ust.1
21. Ustawa o finansach publicznych z dnia 26 listopada 1998 r., Jednolity tekst Dz. U. z 2003 r. nr 15 poz. 148 z późn .zm. art.18 ust.3 pkt.2 ustawy
22. Sherali, H.D., Sivanandan, R., Hobeika, A.G.: A linear programming approach for synthesising origin-destination trip tables from link traffic volumes. Transportation research B 28, 213–234 (1994)
23. Sherali, H.D., Arora, N., Hobeika, A.G.: Parameter optimisation methods for estimating dynamic origin-destination trip-tables. Transportation Research 31 B, 141–157 (1997)
24. `http://www.sas.zmp.poznan.pl`
25. `http://www.scats.com.au`
26. `http://www.adt.pl/prasa_ssr`
27. Woch, J.: Kształtowanie płynności ruchu w gęstych sieciach transportowych, Wyd. Szumacher, Kielce (1997)
28. Van Zuylen, H.J., Willumsen, L.G.: The most likely trip matrix estimated from traffic counts. Transportation Research B 14, 281–293 (1980)
29. Yang, H.: Heuristic algorithms for the bilevel origin-destination matrix estimation algorithm. Trnaportation research B 29, 231–242 (1995)

The Influence of Telematics on the Utility Value of the Public Transportation

Ryszard Janecki and Stanisław Krawiec

Faculty of Transport, Silesian University of Technology
Krasińskiego 8, 40-019 Katowice, Poland
{ryszard.janecki,stanislaw.krawiec}@polsl.pl

Abstract. In the past two decades large changes in the structure of traffic flows were noted, in particular in the urbanised areas. The shift has been towards passenger cars, causing a significant drop of passenger volumes in public transportation. As the increasing congestions effects are noted, the above mentioned tendency must be reversed in the coming years. The pace of the changes depends on the perception of the public transportation in the society. A positive assessment of the usefulness of public transportation is related to its utility value which may be increased by the intensive implementation of telematic technologies.

Keywords: Utility value of a transportation system, public transportation, telematics.

1 The Dynamics of Changes in the Modal Differentiation of the Public Transportation

The perception of the transportation as one of the leading elements of the socio-economic policy is currently characteristic of both developed economies as well as in the emerging markets. A particular role in the development processes is played by the cities, the conurbations and the regions. A necessary element supporting their operation is a public transportation system.

The factors building up the requirements for the transportation systems in the areas of intense social and economic activity are, among others, the current status and the tendencies observed in the distribution of transportation tasks. The Upper Silesian conurbation is an example of such a region.

Over the last two decades the share of particular ways of relocating has changed to the benefit of passenger car transportation (Fig. 1, Fig. 2, Fig. 3 and Fig. 4).

In 1988 at the territory of the Katowice Voivodship (of that time) only 5.1% of travel was with individual transportation (private vehicles) – cf. Fig. 1. and Fig. 4. In 2009 in thirteen largest cities of the Upper Silesian conurbation the share grew to 28.5% of the total volume of travel cf. Fig. 3. and Fig. 4.

One of the main reasons for this effect is the growing dependence of the urban population on passenger cars (Fig. 5) and unattractive if not poor offer or the public transportation sector.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 374–383, 2010.

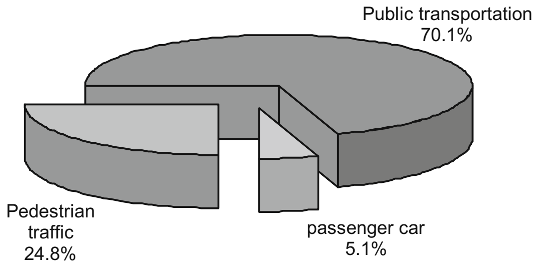


Fig. 1. The distribution of transportation tasks in the Upper Silesian Voivodship in 1988 [8]

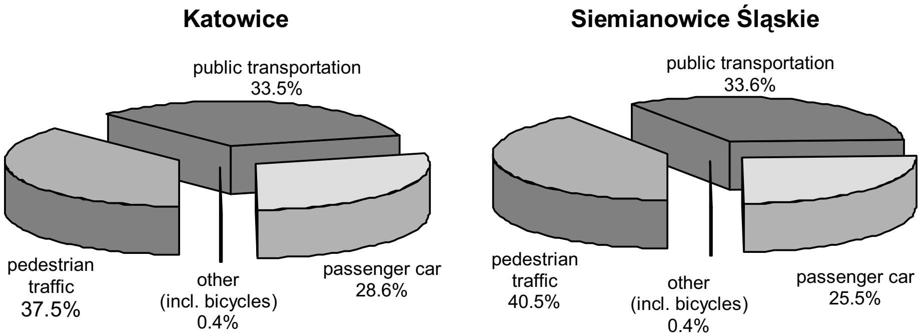


Fig. 2. The distribution of transportation tasks in Katowice and Siemianowice Śl. in 1998 [12]

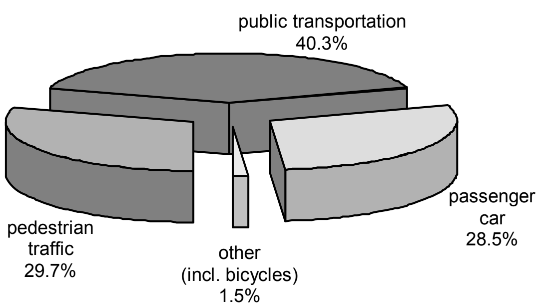


Fig. 3. The distribution of transportation tasks in thirteen largest cities of the Upper Silesian conurbation in 2009 [8]

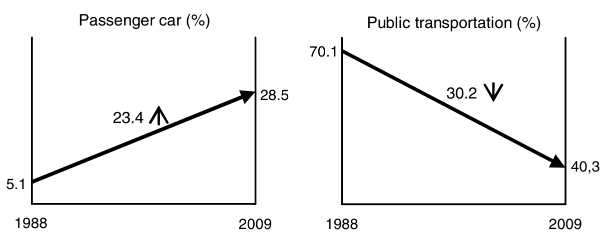


Fig. 4. The changes in modal distribution in years 1988 – 2009 in the Silesian Voivodship [4]

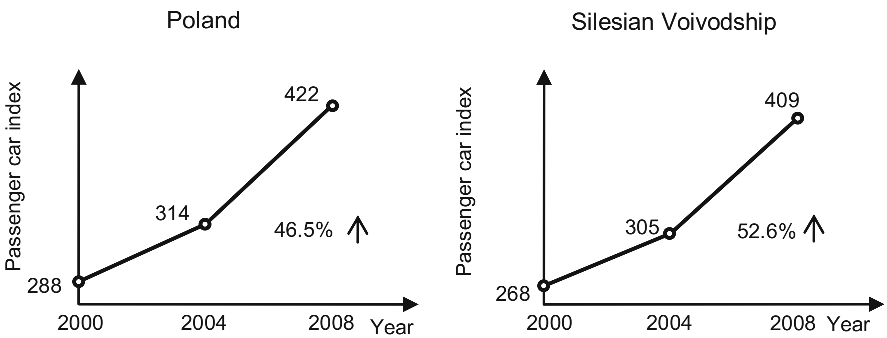


Fig. 5. Passenger car index for Poland and Silesian Voivodship in years 2000 – 2008 [4]

In the situation when private motoring experiences dynamic growth and the possibilities of extending the existing road and street infrastructure are limited, several detrimental effects are noted in the cities, conurbations and in many regions:

- growing traffic load of the road and street infrastructure, reaching values above the limiting capacity
- growing parking space deficit
- the reduction of the transportation volume in public transportation (Fig. 6)
- the growth of the negative influence of the transportation on the environment.

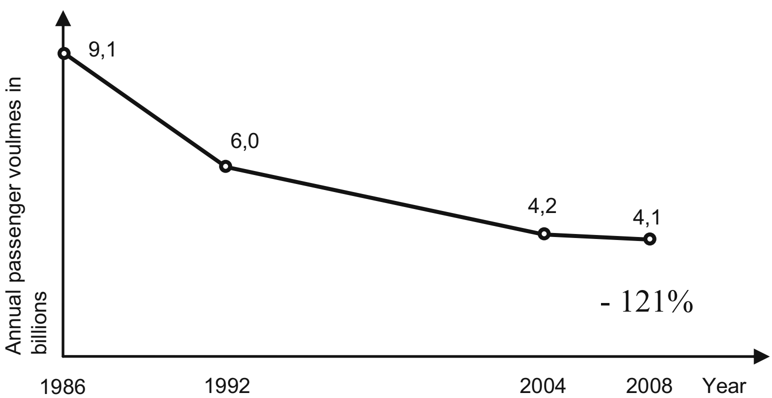


Fig. 6. Transported passenger volumes in public transportation in the cities in Poland, years 1986-2008 [4]

To counter these effects, successful implementations of telematic solutions are necessary, such as intelligent transportation system technologies.

2 ITS Applications in Public Transportation

ITS technologies are currently one of the key instruments for the improvement and the increase of effectiveness of transportation systems, including the public transportation [1], [2], [9], [10] and [16]. They can bring about a step improvement in the transportation service offer, in the system usefulness, its availability and degree of integration. Thus, these technologies significantly improve the potential value of the public transportation systems in the cities, conurbations and regions. Table 1 presents selected ITS applications which may be utilised in the public transportation domain; their degree of complexity has also been indicated. Three-level rating scale has been assumed. The solutions have been classified as individual, sub-complex and systemic.

Table 1. Selected ITS applications for public transportation systems

ITS Module Name	Degree of complexity of the solution
1	2
1. IT tools related to large data volume handling supporting the public transportation management process	Individual solution
2. Innovative IT tools allowing exploratory analysis of the data acquired by the transportation management systems	
3. Location services for public transportation	
4. Fare collection	
5. Passenger stream identification	
6. Cross-charging system	
7. Driver and vehicle schedule planning	
8. Traffic controller support	
9. Transportation services cost calculation	
10. Passenger information systems	
11. Urban public transportation supervision including monitoring systems in vehicles	
12. System for automatic passenger stream measurement	Sub-complex solution
13. Electronic charging system	
14. Public transportation management system	
15. CCTV monitoring system	
16. Passenger information system (external and internal)	
17. Holistic ITS module implementations (pos. 12-16)	Systemic solution
18. Holistic application of extended ITS modules:	
− Systems listed herein (pos. 12-16)	
− Weather system	
− Passenger car user support system	
− Emergency centre system and 112 Emergency Code handling system	
− Motorway management system	
− Emergency incident notification system	

Source: Own research based on [7], [15]

When listing the ITS modules, the degree of complexity (holistic context) was accounted for, as this feature of the module influences strongly the potential value of a transportation system of a specific city, conurbation or a region.

3 The Perception of the Value of Public Transportation by the Society in the Context of the Development of Telematics

Every public transportation system considered may be interpreted as a set of elements and of the relations linking these elements into one functional entity, i.e. as a set of technological, human and organisational means employed in service delivery. The system is therefore characterised by a specific value potential which should grow together with the development of the specific transportation system. Its utilisation and its perception by the society are dependent on:

- the elements forming the system and their widely understood quality,
- internal and external relations linking the elements and the structures of the transportation system,
- the rules and criteria for selection of the elements of the structures, of the relations between them and between them and the environment,
- the activities forming the value adding processes in public transportation,
- the knowledge on the possibilities of creating the innovative mechanisms making possible and conditioning the right selection and linking the elements allowing effective functioning of the public transportation.

A traditional measure of this potential for the society may be the so-called availability, specified in different works with adjectives such as ‘communication, time, territorial, transportational or spatial’. The availability is most often understood as [14]:

- the real possibility of utilising the transportation network; it reflects the degree of transportational development of a specific area,
- the chances or opportunities allowing making use of different kinds of economic and social activity,
- a variety of time and spatial capabilities of different social groups dependent on the age, economic status, family roles and stage in life.

The availability is characterised by a set measures and indicators such as the degree of infrastructure development, travel time, transportation cost, day availability, potential availability, time budgets of individual social categories (the users), time spent at the destination, frequency of the access need, maximum pedestrian access distance (connecting runs), time limit for the journey [11], [13].

The generic form of the availability function (D_{ij}) may be expressed by the formula [3]:

$$D_{ij} = a_m \cdot X_i + X_j \cdot f_m\left(k_{ij}\right), \tag{1}$$

where:

a_m – a constant characterising an intensity to travel to fulfil the motivation m,
X_i – travel destination potential (e.g. a number of inhabitants),

X_j – absorption potential of the destination,
$f_m(k_{ij})$ – a function expressing dependence of the number of travels for the motivation m on the total travel costs (time, distance, price etc.).

Depending on the purpose and the subject of the analysis, the function is used in different forms.

In public transportation the increment of the value potential is achieved to more and more extent by implementing the IT technologies. Telematics, being an area of technology linking the IT and telecommunication technologies, offers currently a wide range of solutions for traffic control and management, also in urban public transportation. This offer, also due to its scope, may eliminate the so-called positive expectations gap effect (Fig. 7).

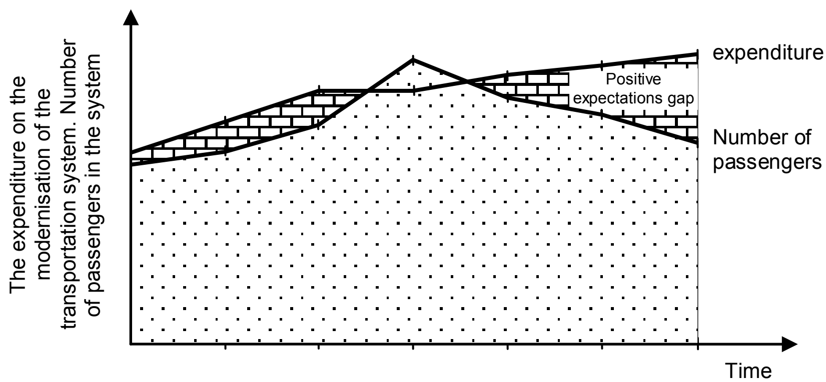


Fig. 7. The positive expectations gap in a public transportation system

The effect appears when the level of expenditure on the modernisation of a public transportation system does not bring about positive changes with the users, such as the increase of the number of passengers using the urban public transportation services [5].

The authors propose therefore a certain measure of the degree of the modernity of a system. The measure can be 'an information availability of the public transportation'. The generic form of this availability function may be:

$$D_{ti} = f\left(s_i, p_d, p_k, \ldots\right). \tag{2}$$

where:
s_i – the telematic infrastructure condition,
p_d – the quality of data acquisition and processing process,
p_k – the quality of mutual communication between the passengers (existing and potential) and the public transportation system.

The condition of telematic infrastructure in a public transportation system depends on the functions and objectives of a specific information subsystem, the expenditure on the maintenance and development of the infrastructure and on the capability to absorb knowledge resulting from the development of data transmission technologies. The quality of the data processing process is related to the creation and utilisation of effective instruments of acquiring, analysis and processing of the information and of the knowledge. It depends on the architecture of a specific subsystem of the transportation system, comprised of the technological platform, databases and data warehouses, right access and development tools as well as the right applications. The quality of the communication of the passengers and the public transportation system is related to the 'structure-behaviour-profitability' triad of this system. The presented parameters of the 'information availability' of a public transportation system have to be suitably disaggregated and presented in a synthetic formula appropriately for each individual system.

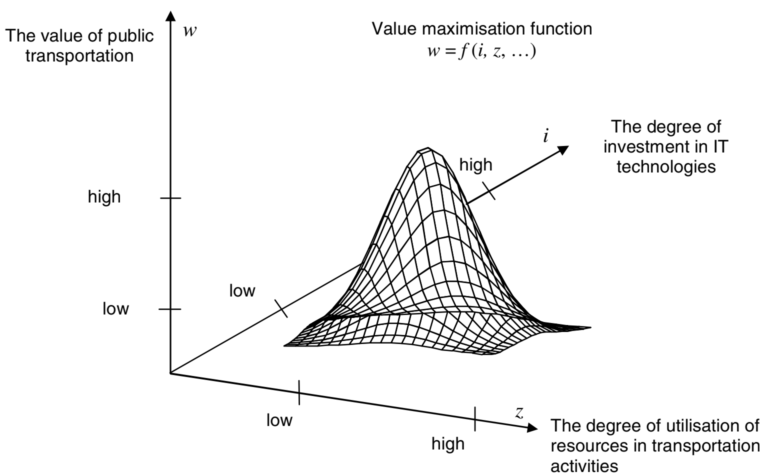


Fig. 8. The value maximisation function for the public transportation

For the public transportation to play a role of a complementary and synergic factor in the network of socio-economic relationships it has to be functional and to maximise its value for the society. In Fig. 8 a graphical representation of the value maximization function for the public transportation is shown. The dependencies include the degree of development of the development projects (including the telematic applications) and the degree of utilisation of resources in the current transportation activities [6].

The degree to which the public transportation meets a variety of expectations of the society (changing in time and space) is a determinant of the value of this system, dependent on the definitions of objectives, their structure and the level at which they are achieved. A specific 'value system' formed by the existing structure of objectives of a functioning public transportation system allows a current comparison of the calculated value and the expected value of the system. The perceived utility value of the public transportation may be considered a measure of its development. The utility value of the public transportation for the society as a function of the expenditure on maintenance and development is shown in Fig. 9.

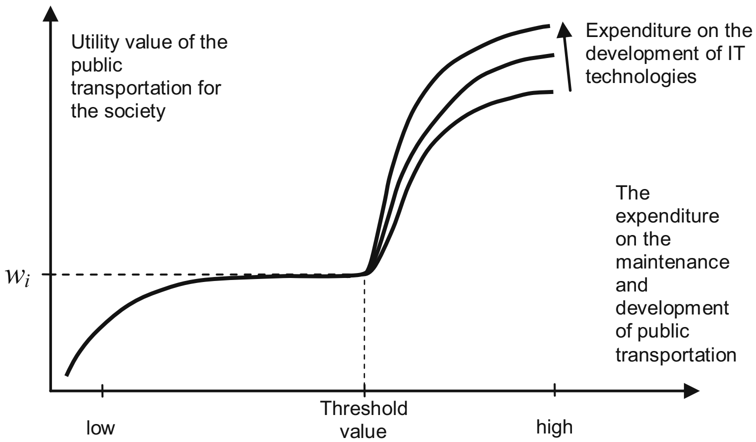


Fig. 9. The utility value of a transportation system as a function of the expenditure on maintenance and development

The threshold value of the expenditure on the maintenance and development of the public transportation is marked on X-axis. After this threshold level is reached, the value of the system for the society grows noticeably. The threshold expenditure allows surpassing the 'incidental value' w_i, i.e. the awareness that the public transportation exists and is operating but its utility is incidental. The increment of the discussed value after the threshold value of expenditure is reached is a function of the expenditure on the development of telematic technologies. That expenditure, currently still at a low level, show a growth tendency. It should not be forgotten, however, that a certain limit of profitability of investment in public transportation telematics exists. After this limiting value is reached, the utility value of the system does not increase any more (Figure 10).

These limiting values are different for specific systems and needs to be studied separately in each individual case.

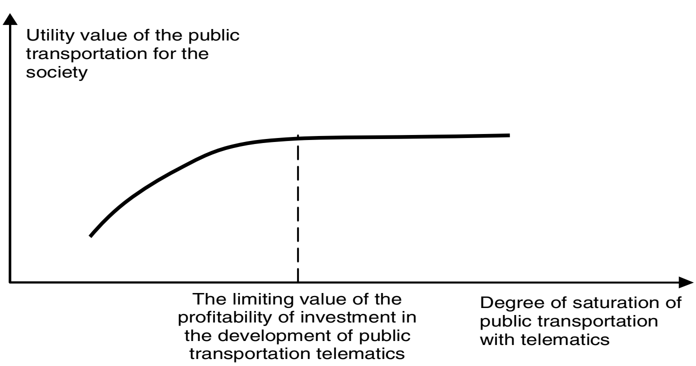


Fig. 10. The increment of the utility value of the public transportation as a function of the expenditure on the development of telematics

4 Conclusions

The development of public transportation should – among others – maximise the value of this system for the society. One of the determinants of such transformation is an information structure, the development of which is conditioned by the development of IT and data transmission technologies. The current paper presents the approach to the study of the influence of telematics on the development of public transportation from the point of view of economic science. The instruments for this research are the analyses of the creation and utilisation of the value potential of the contemporary public transportation. The process of the maximization of public transportation value in model approach has been presented. An information availability function has been defined, which – together with the traditional spatial availability function – may be a measure of development of a public transportation system.

References

1. Bekiaris, E., Nakanishi, J. (eds.): Economic Impact of Intelligent Transportation Systems: Innovation and Case Study. Research in Transportation Economics, vol. 8. Elsevier Ltd Oxford, Amsterdam (2004)
2. Button, J.K., Stough, R.: Telecommunications and Intelligent Transportation Systems. In: Buton, J.K., et al. (eds.) Telecommunications Transportation and Location, pp. 137–142. Edward Elgar Publishing Ltd., Cheltenham (2006)
3. Domańska, A: Wpływ infrastruktury transportu drogowego na rozwój regionalny. p. 49 Wydawnictwo Naukowe PWN, Warszawa (2006)
4. Janecki, R.: Modelowanie podziału zadań przewozowych na obszarze aglomeracji – specyfika problemu. In: Materials of the International Conference 'Structuring Communication Behaviours of the Society by the Promotion of Public Transportation ("Kształtowanie zachowań komunikacyjnych społeczeństwa poprzez promocję transportu zbiorowego"), Wyższa Szkoła Planowania Strategicznego, Dąbrowa Górnicza (2010)
5. Janecki, R., et al.: Warianty koncepcji miejskiej komunikacji autobusowej w śródmieściu Rybnika w kontekście projektu Rybnickie Metro" etap I. R and D Project Report . Faculty of Transportation of the Silesian University of Technology, Katowice, pp. 13–14 (2009)
6. Janecki, R., Krawiec, S.: Influence of Telematics on Transportation System's Value. In: Mikulski, J. (ed.) Advances in Transport Systems Telematics, WKiŁ, Warszawa, pp. 119–121 (2009)
7. Janecki, R., Krawiec, S., Sierpiński, G.: Telematics and the Transportation System's Value. Paper accepted on 2nd IFAC Symposium on Telematics Applications, Politehnica University Timosoara (October 2010)
8. Karoń, G., et al: Studium Wykonalności: Program inwestycyjny rozwoju trakcji szynowej na lata 2008 – 2011. Analiza ruchu. R and D Project Report. Faculty of Transportation of the Silesian University of Technology, Katowice, p. 71. (2009)
9. Koźlak, A.: Inteligentne systemy transportowe jako instrument poprawy efektywności transportu. Logistyka nr 2, płyta CD (2008)
10. Quinet, E., Vickerman, R.: Principles of Transport Economics, pp. 350–353. Edward Elgar Publishing Ltd., Cheltenham (2006)
11. Spiekerman, K., Neubauer, J.: European Accessibility and Peripherality: Concepts, Models and Indicators, Nordregio, Stockholm, pp.11, 16–35 (2002)

12. Starowicz, W., Rudnicki, A., Janecki, R.: Kompleksowe badania ruchu w Katowicach i Siemianowicach Śląskich. Synteza wyników, Monografia z. 6., Wydawnictwo SITK Oddział w Krakowie, Kraków p. 41 (1999)
13. Taylor, Z.: Przestrzenna dostępność miejsc zatrudnienia, kształcenia i usług a codzienna ruchliwość ludności wiejskiej, Wydawnictwo Continuo, Wrocław, pp. 8–11, 34–50 (1999)
14. Tomanek, R.: Funkcjonowanie transportu, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice, p. 35.(2004)
15. Wilczek, M. (ed.): Inteligentny System Zarządzania Transportem Publicznym. Wydział Transportu Politechniki Śląskiej, Katowice (2008)
16. Wydro, K.B.: Telematyka – znaczenie i definicje terminu. Telekomunikacja i Techniki Informacyjne nr 3-4, 25–26 (2008)
17. Mikulski, J.: Contemporary situation in transport systems Telematics. In: Mikulski. J. (Ed) Advances In Transport Systems Telematics, Silesian University of Technology, pp. 1–12, Katowice (2007) ISBN 978-83-917156-6-6

ZEUS Concept and Its Wider European Application

Ryszard Krystek, Joanna Żukowska, and Lech Michalski

Gdansk University of Technology, Faculty of Civil Engineering,
Narutowicza 11/12, 80 – 233 Gdansk
{rkrystek,joanna,michal}@pg.gda.pl

Abstract. The objectives of the ZEUS project and the resulting concept of an integrated system of transport safety are recapitulated in the article. The context for transport safety management that has evolved since this became a concern of the European Union under the Treaty of Union in 1993 is outlined, and some issues related to applying the ZEUS concept across Europe are discussed. It is concluded that there is scope for exploring ways of doing this, but that the balance of interests at the national and European levels and the difference in nature between safety management on the roads and in the other three main modes will need to be recognised.

Keywords: Safety, transport, integration, integrated transport safety system.

1 Introduction

The system of transport incurs enormous losses. In EU countries transport accidents, especially road accidents, are the number one cause of death for external reasons among people aged up to 45 and the total loss is estimated at 200 billion Euros annually which is more than the annual budget of the European Commission [1]. Each year Poland loses about 30 billion PLN, more than 2% of its GDP. In Poland the number of people killed in transport is almost 5 000 annually and injuries are ten times more with 20% of the victims left disabled for the rest of their lives. In Poland's Transport Policy for the Years 2006 – 2025 one of the basic tasks is to "Ensure a safe transport". The Policy identifies development goals and how they should be achieved both in an integrated system and in the particular modes. This leads to the need for effective organisational, technical, legal and financial measures and ways to integrate them to ensure that strategic goals are realised [2].

It is assumed that just like in other spheres of life, safety of transport will also benefit from integration as a result of more effective prevention and rescue efforts. An overview of international experience and domestic transport safety needs suggests that there are at least three areas where integration is required. They are:

- prevention measures carried out by bodies responsible for safety management,
- monitoring of safety and safety information by integrating transport safety databases,
- independent investigations of transport accidents and formulation of recommendations to improve the effectiveness of prevention.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 384–390, 2010.

A comprehensive approach to improved transport safety means going beyond transport and taking advantage of other transport related policies. Integrated transport policies are a possibility because they involve multiple layer integration within transport and integration with the environment and spatial planning across all levels of public administration.

2 Concept of Integrated Transport Safety System in Poland

The need for an integrated approach to many aspects of transport safety legitimises proposals to change the relevant functional and institutional structures in Poland. This problem was addressed in the three year long research project “Integrated Transport Safety System” (called ZEUS). The preliminary objective of the project is a conception of the integration of safety management systems in four modes of transport: road, rail, air and water to build one, consistent, thoroughly researched and legally binding system of transport safety within all levels of functioning: administrative, legal, technical, informative and human resource. The project has identified four areas of integration that can include the following measures:

- prevention: programming of safety improvement, promotion of risk management methods, effectiveness evaluation of safety measures, training of operators and staff, operator licenses,
- monitoring: regularity of terms, integration of accident databases, use of modern monitoring technology,
- accident investigation: procedures and methods of accident investigation, formulation and promotion of recommendations of necessary changes in the safety system,
- rescue: rescue procedures, methods of collaboration, equipment standards.

Nearly all of the areas of integration have to have a development policy to ensure integration with the other modes. In order to integrate the system of transport safety, there must be integration within the modes. This is why ZEUS has dedicated so much effort to developing mode specific integrated systems of safety management. This includes road transport which is clearly the biggest challenge. The system of road transport management is underdeveloped and far from the well organised systems in the other modes.

As regards the institutional structure of an integrated system of transport safety, it has to be said that Poland’s road transport system should be thoroughly redesigned. This would include the development of the existing functions and the appointment of bodies that are already well established in aviation and rail transport. Both modes in Poland have their own transport authorities and accident investigation commissions. For the integration to continue the entire safety system must be brought up to the same standards. Ultimately, the safety management system should be based on four key organisations (Fig.1):

- for integrated transport safety management - Transportation Office,
- for integrated transport safety monitoring - National Transport Safety Observatory,
- for integrated transport accident investigation - National Transport Safety Board,
- for co-ordination – National Transport Safety Council.

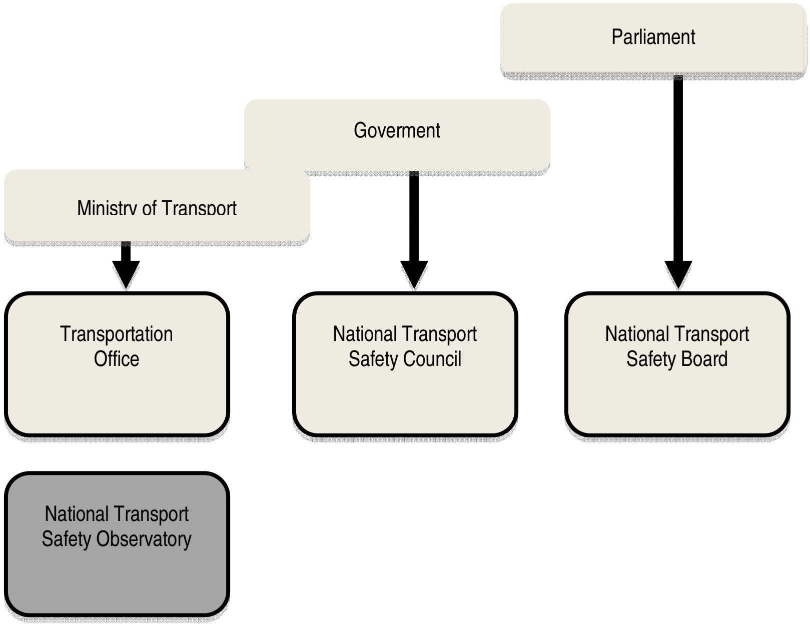


Fig. 1. ZEUS' institutional vision of an integrated system of transport safety

The remit of each of these organisations will cover four modes of transport. For this vision to become reality, the modal safety systems as they are today will have to be modified. The change will involve additions, reorganisation or the appointment of new structures.

Apart from the formulation of new tasks and competences of the new organisations, new laws will have to be adopted which may mean a staged process of transformation. How the road safety management system should adapt to the integrated transport safety system is one of the areas and challenges covered by the ZEUS project.

3 Competences of the Main Organizations

3.1 Transportation Office

The objective of the integrated Transportation Office (or, in the first stage, four cooperating mode specific agencies) is to regulate the transport market and oversee transport safety. The agency will primarily:

- develop strategies, programmes, safety plans,
- oversee transport, the operation of infrastructure, qualifications of staff, training and exams;
- give certification for facilities, products, processes, services, technologies,
- give competency accreditation,
- issue licenses for specific activities,

- authorise the operation of infrastructure and transport activity,
- control compliance,
- carry out periodical statistical safety analyses and evaluations,
- run databanks and registers,
- initiate and evaluate new transport regulations.

The integrated Transportation Office could be established by transforming and adapting Poland's current agencies responsible for rail, aviation and maritime and setting up a new road transport agency (e.g. using the Road Transport Inspectorate as the basis and extending its powers). The Swedish Transport Inspectorate could serve as a model of this kind of structure.

3.2 National Transport Safety Observatory

The objectives of the Observatory are to inform about transport safety, transport accidents, level of safety and how it changes, safety analysis results and to provide safety know-how. It will primarily:

- collect safety information,
- maintain an integrated accident database,
- analyse accident data
- monitor changes in the epidemiological factors, safety measures and their effectiveness,
- prepare and disseminate transport safety information, inform the public about changing risks.

The Observatory could be established by extending the activities of the Polish Road Safety Observatory that is just being built in Motor Transport Institute as an element of the European Road Safety Observatory (ERSO).

3.3 National Transport Safety Board

An inter-departmental advisory body for the government on transport safety. It should have a Secretariat with executive powers which is part of the Ministry of Infrastructure.

The objective of the Board is to identify the government's transport safety policies and coordinate the relevant work. The Board will:

- propose government transport safety policies,
- coordinate the work of institutions within the department of transport and other departments,
- propose areas of research,
- initiate and evaluate new laws in the area of transport.

The Board will be required to submit annual transport safety reports and reports on the work of the Transportation Office, National Transport Safety Council, National Transport Safety Observatory and the other stakeholders within transport safety system. The reports will be presented every year to Parliament and Senate.

3.4 National Transport Safety Council

The objective of the Commission is to provide an independent and impartial identification of the causes of transport accidents. It will:

- carry out in-depth investigations into the causes of transport accidents,
- carry out special analyses of safety problems,
- evaluate the effectiveness of prevention by the government's transport agencies,
- evaluate the effectiveness of rescue and care for victims of transport accidents,
- formulate recommendations on the necessary changes of law and improvements in how transport safety is managed.

The Commission can be established by merging the existing aviation accident commission and rail accident commission and the proposed maritime accident commission with experts on road accident investigation. America's National Transport Safety Board and Dutch Safety Agency could be the models.

4 The European Perspective for Integration of Transport Safety

Under the Treaty of European Union in 1993 a general duty in respect of transport safety was recognised in European legislation, and the European Union (EU) was empowered to act in this area whenever its actions could be shown to give added value over and above what Member States can achieve individually – *i.e.* consistently with the principle of subsidiarity. This has led to the establishment of the European Aviation Safety Agency (EASA), the European Maritime Safety Agency (EMSA), the European Railway Agency (ERA), whose remit includes safety, and a Road Safety Unit in what is now the Directorate-General Mobility and Transport (DG-MOVE). Any move towards implementation of the ZEUS concept across the EU would need to take these organisations and their *modus operandi* as its starting point.

The Road Safety Unit has adopted and pursued a series of Road Safety Action Programmes (RSAP) and in 2001 the EU set an aspirational target to halve the annual number of deaths in road accidents in the EU by 2010. Although the reduction achieved over this period will fall short of the target it has been substantial, and the next RSAP for 2011-2020 is awaited with interest. A framework research project known as SafetyNet has laid scientific foundations for improved collection and analysis of data related to road safety and established the European Road Safety Observatory (ERSO). In the meantime the independent non-governmental European Road Safety Council (ETSC) has been monitoring road safety across Europe using existing data through its road safety performance index programme PIN.

In 2003, the European Commission decided to appoint a group of experts to advise on a strategy to deal with accidents in the transport sector, including a strategy for safety investigation. The group reported [3] with an overview of issues relating to the main means of transport leading to mode-specific recommendations and a number of general recommendations. Early in its work the group identified the need to provide guidance on the methodology for safety investigations and appointed a methodology working group, whose work led to a separate report providing the group of experts' recommendations concerning accident investigation in the form of a set of guidelines [4]. This must also be taken into consideration when it comes to the wilder application of the ZEUS concept.

In relation to both aviation and marine accidents the group identified tension between the requirements of technical safety investigation with a view to learning lessons and preventing future accidents and those of more judicial investigations involving apportionment of blame. Practice concerning the publication of reports on technical safety investigations was less than consistent. In relation to rail accidents, the group made many recommendations concerning the implementation of Directive 04/49 on the safety of the EU's railways and noted a movement towards common methods of investigation. The group also considered pipelines and noted that self-regulation with the help of pipeline integrity management systems seems to be working well, except perhaps in respect of third party interference [5]. More generally, the group commended its guidelines on methodology, identified the need for co-ordination across Europe among bodies responsible for safety investigation, for common databases, for common training standards for investigators, and for clarification of the role of the three European transport safety agencies.

The guidelines on methodology call for Member States to designate Safety Investigation Authorities empowered and resourced to conduct independent effective and competent investigations of accidents in all transport sectors for the improvement of transport safety. These authorities should be guaranteed independence, especially from those responsible for the establishment and enforcement of safety requirements. Technical safety investigations should have priority over judiciary investigations unless there is clear evidence of a serious criminal act. Confidentiality of evidence given to safety investigations should be protected, even in the context of other investigations of the same accident. Reports of investigations should be addressed to those responsible for implementing the recommendations made, and both the reports and the actions taken in response to them should be published.

In relation to the lack of systematic technical safety investigation of road accidents, the SafetyNet project has subsequently produced recommendations [6] for their transparent and independent investigation on a European basis. These address the issue of sampling, having regard to the need for an expert team to go quickly to the site of each accident sampled for investigation as soon as practicable after the accident becomes known to the police or highway authority. They also recognize that sampling will need to reflect national requirements for the findings from the investigations as well as any agreed pan-European requirements.

5 Conclusions

It is thus recognized at the EU level that there is great scope for improvement in the management of transport safety across Europe, including wide variations in practice within Member States, with the excellent examples of The Netherlands (Dutch Safety Board) and Sweden (Transport Agency), which have influenced the thinking of ZEUS so strongly, being the exception rather than the rule. Moreover, there is a good deal of convergence between the ZEUS concept and the recommendations of the group of experts who reported to the European Commission in 2003.

Scope certainly exists, therefore, for exploring ways of applying the ZEUS concept across the EU, but this will need to be done in ways that recognize the delicacy of the balance between mutual interest in achieving successful working of EASA, EMSA and ERA and the individual interests of the Member States and their national transport safety organizations. It will also need to recognize the complexity of achieving consistency and integration of safety management on the one hand in the three modes in which operation is largely professionally organized and managed and accidents are few enough for all serious ones to be investigated in depth, and on the other hand in the road transport system in which every citizen is a responsible participant, much of the use of the system is by individuals free to go about their own activities in their own ways, and accidents are far too numerous for more than a tiny proportion to be investigated in depth.

References

1. White Paper — European transport policy for 2010: time to decide. Office for Official Publications of the European Communities, Luxembourg (2001)
2. Krystek, R., Sitarz, M., Zurek, J., Gucma, S.: Integrated Transport Safety System – ZEUS. In: International Conference on Safety and Reliability Systems KONBiN, Wroclaw (2008)
3. European Commission: Final report of the group of experts to advise the Commission on a strategy to deal with accidents in the transport sector 2004-2006, Brussels (2006a)
4. European Commission: European methodology for safety investigation of accidents and incidents in the transport sectorM Brussels (2006b)
5. Allsop, R.: Integrated Transport Safety System. In: Final Conference of ZEUS Project. Gdansk University of Technology, Gdansk (2010)
6. SafetyNET: Recommendations for transparent and independent road accident investigation. Deliverable D4.5 of the EU FP6 project SafetyNet. Directorate-General Transport and Energy, Brussels (2008)

Telematic Problems of Unmanned Vehicles Positioning at Container Terminals and Warehouses

Stanisław Kwasniowski, Mateusz Zajac, and Paweł Zajac

Wroclaw University of Technology, Faculty of Mechanical Engineering,
5 Łukasiewicza St, 50371 Wrocław, Poland
{stanislaw.kwasniowski,mateusz.zajac,pawel.zajac}@pwr.wroc.pl

Abstract. This paper describes the issues of transshipment container terminals operations, in the light of the development of this kind of transport. An increase in handling requires an expansion of stacking yard and automation of handling and transport processes. The development in this area first and foremost depends on modern handling technologies and automatic identification systems. AGV trucks play a key role in in those systems. The role of universities is to promote innovative technologies. Paper [2] contains the status of intermodal terminals development in Poland, which was awarded the prize of the Minister of Infrastructure of Poland in the field of "organization and management." The paper contains a detailed description of the principles of positioning, control and propulsion of AGV vehicles. The content was developed to make it understandable to logisticians responsible for the implementation question in Poland[1].

Keywords: AGV, container terminals, logistics.

1 Introduction

These days, on the basis of container transport, we are observing an extraordinary development of intermodal transport. In some countries the container loads account for 30% of the transport weight. According to many sources [1], [3], [4] and [7], in countries of the European Union the container transport accounts for, on average, 15% of loads and it is rising at the annual rate of 14%. The number of container loads in Poland is much lower – 3.5% in 2008 - but the growth rate is higher and amounts to 20%. An increase in shipments forces the intensification of handling and transport, which in turn force using automated equipment and automated vehicles, which are telematically controlled.

The essential issue is a computer management of container depots and great handling and transport activities.

Typical examples of this tendency are the biggest European container terminals in Rotterdam and Hamburg. Processes of loading and unloading, moving and storage are controlled by a computer.

---

[1] The authors conducted studies using the world's published literature databases: EBSCO, COMPENDEX, PROQUEST, BAZTECH and e-Journals. The results of the analysis is not given in the paper because of space needed, will be published in a separate monograph.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 391–399, 2010.

Diverse ranges of machines and devices are used at containers reloading in intermodal terminals. A constantly growing, year-on-year, stream of containers leads to the situation, where the handling techniques, which have been prevalent so far, become insufficient. This equipment comprises harbor gantry cranes, terminal cranes, straddle carriers, fork lift trucks to transship empty containers, reloading trucks, AGV and ALV trucks, self propelled unmanned goods wagons.

This restriction does not concern the design abilities and technology. It is related to the precision of an operator (determined by the diameter and distance between slots, in the corners of containers) and labour efficiency. The above-mentioned machines are taking part in logistic activities like: containers unloading from means of transport (sea boats, inland vessels, railway cars, and motor trucks), loading means of transport, handling the containers within horizontal transport and vertical transport, which are determined by distances within terminals.

AGV vehicles are known in the world - also in Poland, their application in the vast majority comes down to production logistics systems - ESP. They are not used for intermodal terminals in Poland [2]. The work of known global positioning systems is characterized by casually, control and propulsion of AGV vehicles. On this basis it is difficult to imagine constructing innovation - prototypes in Polish conditions. The paper is aimed at presenting the authors achievements in this field.

2 AGV Vehicles

Automated Guided Vehicles (www.gottwald.de or [2]) are able to work in two versions of engine solutions: combination of a Diesel and hydraulic drive or a Diesel and electric drive. Combustion engine's power amounts to 350 HP. The maximum speed amounts to 21 km/h.

Table 1. Comparison of technical parameters of AGV container vehicles [5], [6]

		TTS Marine	Gottwald
Sort of load	Container 40'	x	x
	Container 20'	x	x
	Container 2x20	x	x
	Containers 30 i 45	x	optionally
Characteristics of load	Maximal weight of the container		40t
	Maximal weight of two containers	92,5t	60t
Size	Length	12,19 m	14,8 m
	Width	2,6 m	3,0 m
	Weight	16 t	25 t
	Size of tyres	225/75 R15	R25
Speed	Maximal speed when driving forward	5,5 m/s	6 m/s
	Maximal speed when driving around a curve	3,3 km/h	3 m/s
Precision of positioning			+/- 3cm
Fuel tank capacity			1200 l

In comparison with traditional attitude toward management of vehicles and storage facilities, at the terminal, which is equipped with AGV trucks the number of vehicles may be reduced by 50%. As a matter of fact trucks are constantly moving.

At large container terminals, where activities like storage, transport inside the terminal and reloading are taking place, a system of AGV vehicles is able to cooperate with fully automated cranes in the system of ASC (Automated Stacking Cranes). Table 1 shows the comparison of AGV vehicles of different makers.

Like the FMS (Flexible Manufacturing System), Flexible Distribution Systems (FDS) make use of AGV trucks, which feature better flexibility, better utilization of workspace, safety and lower costs of operating in relation to systems, which use different forms of terminal transport.

For the first time on a large scale, AGV trucks were used in 1974 in a Volvo factory in Sweden. Over 10 years later, 3300 companies were using 15000 AGV trucks. In Canada, there is the biggest manufacturing system, which is a part of the General Motors Company. This system uses more than 1000 AGV trucks. Also the New York Times uses trucks to transport newspapers from the printing house to the store-room. AGV trucks are also used to transport luggages at airports. But Japan is without doubt a leader in terms of using AGV trucks. In 1989, Japanese enterprises bought 5000 AGV trucks, while in Europe the number of trucks that were bought amounted to 3000, and what is more, 500 trucks were bought in the United States.

In terms of trucks' working mode, two kinds of working mode could be found in literature:

1. 'stop and go' mode,
2. 'pick and drop' mode.

A truck that works in the first mode has to execute a concrete task during being in the system. During a transport operation, the truck moves in the direction of given crane, then it waits for the end of a reloading operation. Before the end of a reloading operation, the truck cannot start a new transport operation. That is why the second working mode is more practical. The truck, that works in the second mode, leaves a transport container in the intermediate storage area (or next to the different gantry crane), then it performs an operation called "empty trip" in the direction of another gantry crane (intermediate storage area) and takes a different container from there, which means that it has started the next transport operation of a different task.

To avoid transport collisions, the system imposes unidirectional taskflow while trucks have to work periodically, and their working mode has to be unidirectional. During individual working cycles of a truck, there is a task which is fed into the system by a loading station, and a task which leaves the system by an unloading station.

At first, the most important element of AVG system was an electromagnetic field detector, which was designed to work in combination with the wire, which was placed around the surface, where the vehicle moves (so it marked a workspace). Essentially the system worked like a transformer that utilized a primary winding, which was placed on an AGV truck to produce a low voltage signal in the wire (placed around the surface) and a secondary winding, placed on the main part of the vehicle to determine the distance from the wire. Thanks to the hodometer and the compass, the system was able to determine distances of moving. At the same time, every obstacle such as people could be noticed by a proximity detector and scanning sensor. Signals, which came from detectors, were fed into a microcontroller, which sent command signals to the driving system and to supportive engines, one per each drive wheel.

The most difficult problem of then designers was a formulation of control application as it required plenty of time to write it, to carry out many-hour tests and to collect data about every possible shape of workspace. It turned out that a vehicle, which did not use an exterior radio transmitter, had to move very precisely. Moving on 80% of terminal area was not a problem. The problem was with moving on 100% of container terminal area, irrespective of a shape of terminal.

At present, laser scanners are used and they combine principles of laser and radar operation. A mirror, which is rotating with a speed of 8 Hz, reflects laser rays (1st class, infrared), which sweeps 300° of space. Every object with the minimal reflection coefficient 1.8 (black…) can be identified within a radius of 6 m. Two levels of security can be chosen, and they could be set up by two zones of any shape:

- "alarm zone", within a radius of 10 meters from device,
- "safety zone", within a radius of 6 meters from device,
- These zones are connected with two independent relay outputs, and this enables creating complex applications:
  - The alarm zone makes it possible to send a light signal or an audio signal, while someone gets in the zone, and this signal means that this person is close to a danger. It allows that person to move back outside the zone, without unnecessary stoppage of the machine,
  - The safety zone is used to stop the machine immediately. It is restarted always after "clearing" the zone. If a manual restart is necessary, it is important to use an additional module of security. This system is unique because of a high resolution (0.5°) and excellent precision and coverage of large area (260m2) at the same time. The scanner meets a standard EN 61496-3, which is binding in the UE.

Instead of designating the workspace by a copper wire, which could be difficult on terminals, the transponders started to be used.

Vehicles are controlled by software, which is delivered with vehicles, and positioning is taking place on the basis of transponders, which are placed on the surface of the terminal, and thanks to this solution, the precision of positioning amounts to about 25 mm.

Until recently, permanent points of reference were used – markers (2nd point of article) and AGV trucks were positioned according to them. However, it was a very complicated and expensive solution. Then GPS satellite systems and more precise DGPS started to be used, but they were appropriate only for large robots, where the precision was not so important.

3 The Issue of Positioning, Navigation and Operation of AGV Trucks at Container Terminals

An ultrasonic sonar, a proximity detector which is an infrared device, a tactile sensor and infrequently used low-powered microwave radar – these devices are used with the aim to secure a sense of AGV truck direction. Because of their simple design and imperfection of physical recognition of the ground, devices are very often connected.

Every single sensor has a different operating range, and it is situated in a way, that a given point of the nearest surroundings is always identified by two of them. The image received is a good reflection of real surroundings. Ultrasonic sonars are the devices used most commonly to measure the distance. This is because of the availability of necessary electronic elements and ease of data processing, which allows constructing maps of surroundings of an AGV terminal truck. New AGV trucks “recognize” surroundings analysing the image that comes from a camera. The imperfection of image recognition technique and a high design power of computers are big obstacles to development of this data collecting method.

First of all a programmer has to create an algorithm, which will be able to present the data graphically or numerically, according to truck’s way of perception. Then with the help of measurements, the truck creates reference points and thanks to them it can change its position. That is why the data is divided into local maps or global maps. The data from local maps is the information about the last measurement, while global maps present the collection of information that could be found several times in local maps.

An AGV terminal truck with autonomous navigation consists of a mobile platform and a computer that supervises the work of robot. These devices are combined by radiomodem through a serial line RS232.

Sending-receiving units are used in the communication module. They enable using bilateral and half-duplex data transmission at a speed of 1200 bits/second. The device works on an allowed 433 MHz band and a power of transmitter amounts to 6 mW, and that is why it is unnecessary to apply for a license. A two-way converter of RS-232 standard is used in a signal of sender-receiver.

Especially designed digital rangefinder (placed on the shaft of stepper motor) is also used here. This enables measuring a distance to the nearest obstacles, round the mobile platform of AGV truck. This layout data, that includes information about surroundings, is transmitted through the radio to the computer, and then it is presented graphically. Piezoelectric elements (that are used in car alarms) were used in the sonar. Microprocessor manages the stepper motor, through the system of feeder and controllers.

On the mainboard there is a microprocessor network that is based on AT89Cx051 systems with 11.059 MHz clock. It is a master-slave type network, where the master is a computer which supervises the work of the robot, and the slave is every microcontroller on the mainboard. Microprocessors have an access to the data bus, which controls the sending-receiving module through the built-in port. The data bus is built on the basis of logic CMOS circuits. Transports were used to raise the voltage, which is necessary to control the radiomodem.

Two built-in sixteen-bit counters (that co-operate with an optical generator of impulses of driven distance) secure the monitoring of positions. A bumper with contact sensors protects from the obstacles that were not noticed by the digital sonar. There is a diaphragm mounted on the axis of engine, on a path of light beam, from the transmitter to the receiver. The transmitter is a diode, which emits infrared, that is modulated with frequency of 38 kHz. Self-contained TFMS 5380 infrared receiver is sensitive to light, modulated in this way. Every break of light beam by the diaphragm causes a short change of condition in the transmitter (from high to low). It happens two times, during one rotation.

In the microprocessor, there are two subprograms which control engines in different ways. The first realizes the remote-control function of a traveling AGV truck. An operator optionally controls the traveling by pressing appropriate buttons on a keyboard of the computer, which results in constant giving orders of traveling in a specific direction. This part of program is used to move a mobile platform to the place where the operation starts or to the parking place. Additionally, impact sensors are also active. If any of them is actuated the remote control is suspended, the robot is stopped and moved 50 mm back. After 1 second the remote control is restored. In this subprogram there is a one-way radio communication and the computer does not receive any data from the mobile platform.

The second subprogram is designed for autonomous control of the robot. The device follows orders of specific parameters. These parameters contain the information about particular distance or angle of rotation displacement. Information is presented as a number of impulses, which are sent to the counters of microprocessor. One unit of path corresponds to 1.4563 mm, and one unit of angle of rotation corresponds to 0.7258°. The program of microcontroller supervises the driving straight ahead where odometer readings of engines have to be equal. Additionally, there is a procedure in the program, which makes it possible to bring engines to a stop very softly. The microcontroller operates engines by a module of feeder with controllers.

After the end of movement the mobile platform sends an acknowledgement of executing a task to the computer. The form of acknowledgement is the same as it gets earlier but it contains zero bites. This mode of control stimulates the activity of buffer, but it produces a different effect. After the crash, the procedure takes over the control, which moves the robot back, to the place where the previous movement has started. Then it sends information about an obstacle and about its location. The obstacle location is recorded in the report, in the form of bites.

The navigation subprogram controls all modules of an AGV truck, according to appropriate algorithm. It is a cyclical process, which is held in following temporary sequence:

- The command of the examination of the ground is sent thought RS232 connection and then through the radiomodem to the sonar,
- The radiomodem, which is installed on the platform of a truck, receives a signal, and then, the signal is sent to the microprocessor network,
- Every processor analyses the message and decides, if this message is addressed to any of them,
- Only the microprocessor, which is responsible for the service of sonar, reacts to the signal and starts measurements,
- Measurements consist in positioning of the sonar head at an angle of n = 3°36' (n $\in$ {0, l, ... , 99}) and determining the distance between the AGV truck and the nearest obstacle,
- The result of measurement is immediately transmitted to the computer, where it is recorded in the form of a sixteen-bit number vector,
- Subsequently the vector is converted into the distance in centimeters and presented graphically, where one centimeter correspond to one pixel,
- When local map is ready, we can check in which of eight directions the mobile platform of the robot can be moved,

– Then, there is an analysis of the algorithm of ground exploration, and next on the basis of analysis, the movement and the direction of the AGV truck is marked, Fig.1,
– The layout data is transferred and saved on a global map, and then it is presented graphically as a collection of measurements on the monitor of the computer,
– The movement of the truck, that was marked earlier, is turned into a command and in this form it is sent to the truck.

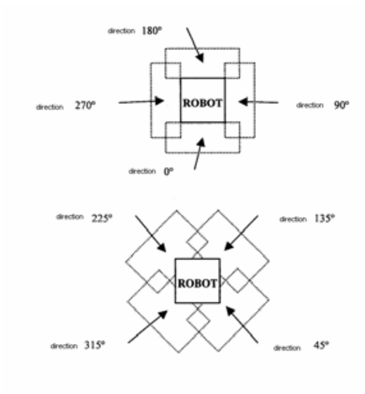


Fig. 1. The way of determining possibilities of movement of AGV truck in given direction. [1], [3]

This cycle is repeated time and again, which leads to receiving of the storage place profile. The algorithm of ground exploration consists of two parts. The first part (which works directly after activation of a truck) sets out a “driving forward”, until the truck encounters an obstacle. Then the computer saves in the memory that there was a contact of the truck with an obstacle, for example with a container or with a fence of a yard, and the truck is addressed to a different place.

The control computer decides to examine the ground around the truck. So the computer sends an appropriate report to the microprocessor, which controls the sonar. After receiving the command, the stepper motor positions the sonar head in the initial position. Then there is a standstill, which lasts for 900 ms and serves to dampen vibrations of the sonar structure, which are disrupting work. Next, the microprocessor turns the generator on (800 ms), which correspond to the length of 25 wave periods of the ultrasonic wave which is sent. After 1.5 ms, there is an activation of the receiver, and the output of this receiver is connected to the microcontroller. The program, which is in the microcontroller, waits for 70 ms for the signal from the receiver. The receiver detects the reflected ultrasonic signal and sets the high level on the output of the TBA 2800 system. The microcontroller responds to this signal and stops the countdown, saves the data in the form of a report and sends it to the computer. If the receiver does not detect an ultrasonic signal, sixteen-bit data is replaced by FFH bites, which means an overflow of the counter. The report contains information about the angle of rotation of the stepper motor. Measurements are repeated 100 times, but every time the angle of rotation is increasing about 3°36'. The angle of rotation and the distance are known, so the computer that supervises the work of the robot, sketches a virtual image of space of the robot, on a monitor of the computer. After this procedure the head moves back to the previous position. The exploration of ground consists in moving along trajectories, at a distance of 10 cm to 40 cm from the line of obstacles. The navigation of the terminal trucks consists only in analyzing the data, which comes from the sonar.

Until now, we can find the following types of examination in the literature: the recognition of surroundings by an ultrasonic sonar, touch of an obstacle (which was not detected by the sonar) by the bumper, range of radiomodem action.

4 Intelligent Rail-Vehicles

4.1 Rail Mounted Gantry Crane

It is a group of vehicles that drive on a truck, and either rail mounted gantry crane and rail vehicles belong to this group of vehicles. In recent years, gantry cranes have revolutionized the way of containers service. Changes that have happened are connected with:

- The crane capacity,
- The number of containers that are serviced,
- Low cost of energy in the container transport,
- Fast and precise automatic service.

Five out of ten biggest container terminals in the world – including international terminals such as CTA (Hamburg), Euromax (Rotterdam), Busan New Port (Korea), Kaohsiung and Taipei (both in Taiwan) – are equipped with automatic crane systems of ABB production.

Every day, over 10 thousands of containers are on average transported in an advanced container terminal. Every relocation of container has to be done very quickly and precisely to the destination. At the same time, it cannot interfere with the transport of other containers and with the work of other hosting cranes and vehicles.

Thanks to the use of AGV technique, one operator, who works in one control room, services 12 cranes. While each of the conventional gantry cranes needs one operator, the application of this solution makes it possible to employ only one person, who supervises the automated monitoring of 8-12 cranes. Monitoring is an automated process and it takes place in the central control room. It leads to reduction of personnel and to the increase in the capacity of the operator's work.

Fully electrical hoisting crane is also very energy-saving. It is because of systems of energy recovery during the braking, driving and lowering.

4.2 Cargo Mover Rail Vehicles

Cargo Mover is a rail variety of an AGV vehicle, but now there are also tests of implementation of AGV into driving of passenger trains [2].

Torsten Dellmann from RWTH Achen, formulated a prototype of Cargo Mover vehicle. A rail car, which is a type of container platform, is equipped with a combustion engine. The vehicle does not have a control cabin, but it is equipped with a camera, which is turned toward the track, in front of the vehicle. Signal that comes from the camera is sent to the traffic management centre. The vehicle could be stopped, if this action is based on the decision, which is made based on the observation of the area in front of the vehicle. The purpose of this vehicle is to transport container loads, from private railway siding and from the zone of a logistic centre, which is in given area. The dispatcher is able to supervise the work of several self-propelled rail cars at the same time. The Cargo Mover is presented on the picture in [2] and www.schleifkottenbahn.de.

5 Conclusions

- The development of intermodal transport and the use of containers leads to the situation, where logistic operators have to use more and more container unit loads and where the processes of management of transshipment and transport control and their optimization have to be automated.
- One of the best solutions is the use of intelligent, unmanned means of transport called AGV trucks and their next generations, which are all computer controlled.
- In these processes, a very important issue is the movement precision, and there are many technical problems connected with it.
- For positioning, different types of duplicated scanning systems and systems of precise localization of mobile means of transport are used.

References

1. Kim, K.H.: Container Terminals and Cargo Systems. Springer, Heidelberg (2007)
2. Kwasniowski, S., Nowakowski, T., Zajac, M.: Intermodal transport in the logistics chain. ublisher of Wroclaw University of Technology, Wroclaw (2008)
3. Korzen, Z.: The logistics systems of internal transport in a warehouse and storage. IliM, Poznan (1998)
4. Dudzinski, P., Madejski, W., Ross, H.J.: SYSTEMS Journal of Transdyscyplinary Systems Science. In: Commercial vehicles with an automatic reversing system and loading of containers at terminals logistics, vol. 8(1), Publisher of Wroclaw University of Technology, Wroclaw (2003)
5. http://www.e-automatyka.eu
6. http://www.mtssensors.com
7. Neider, J., Marciniak-Neider, D.: Intermodal transport. PWE, Warszawa (1998)

Estimation of Socio-Economic Impacts Using a Combination of Fuzzy-Linguistic Approximation and Micro-Simulation Models

Tomas Starek[1] and Miroslav Svitek[2]

[1] Telematix Services, a.s.,
Na Žertvách 34, 180 00 Prague 8, Czech Republic
starek@telematix.cz
[2] Czech Technical University in Prague, Faculty of Transportation Sciences,
Dept. of Control and Telematics,
Konviktská 20, 110 00 Prague 1, Czech Republic
svitek@fd.cvut.cz

Abstract. The paper focuses on a detailed introduction of innovative socio-economic and qualitative ITS (Intelligent Transport Systems) impacts' estimation approach which also allows the transparent calculation of the transportation external costs. The methodology proposed utilizes the mathematical background of the fuzzy-linguistic approximation that in the case of insufficient expert knowledge-base is combined with the transport micro-simulation models' outputs.

Keywords: ITS evaluation, fuzzy-linguistic approximation, micro-simulation, socio-economic impacts, effectiveness.

1 Introduction

The paper introduces the approach leading towards the elimination of the key absence of the contemporary ITS evaluation methodologies. It has been represented by insufficient methodology steps on how to deal with the incorporation of socio-economic and qualitative impacts evaluation. As these often represent key background for the determination of the final recommendations related to the implementation of particular transport-telematics application it is of high importance to cover them in a uniform and transferable way within the complex methodology.

A potentially suitable mathematical apparatus has to be chosen as the starting point. For this purpose the research on possible deployment of Kalman filters, neural networks and fuzzy-linguistic approximation has been performed. In the first two analyzed math constructs the need for long time-line intervals related data was identified, which represents one of primary obstacles in the area of intelligent transport systems' evaluation. The input information required is not reachable at all or in the short time-line periods only. As the fuzzy-linguistic approximation methodology does not require this kind of inputs it is

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 400–409, 2010.

the right candidate for an in-depth research of its potential related to the above mentioned purposes. The following chapters will therefore focus on the methodology introduction utilizing the fuzzy-linguistic approximation as a tool for the socio-economic and qualitative impacts estimation in the area of intelligent transport systems.

1.1 Fuzzy System Background

In spite of presumptions, the knowledge of the fuzzy-linguistic approximation principles and its theoretical background is not further presented, from the perspective of connecting considerations it is proper to mention its basic set up - fuzzy system selection.

The theoretical essence of the Mamdani and Sugeno fuzzy system is defined below. The basis for the description of both types is represented by the following function of several variables [3].

$$y = f(x_1, x_2, \ldots x_n) \tag{1}$$

In the case of Mamdani fuzzy system the output variable y is defined on the universe Y and the input variable x_i on the universes X_i. If universe Y is covered by fuzzy sets B_j and universes X_i by fuzzy sets A_i^j then a non-linear function f can be approximated by fuzzy system represented by r rules of the following type [3].

$$\mathit{If}\ \left(x_1 = A_1^{jk}\right)\ \mathit{and} \ldots \mathit{and}\ \left(x_n = A_n^{jk}\right)\ \mathit{then}\ \left(y = B^{jk}\right) \qquad k = 1, 2, \ldots, r \tag{2}$$

This fuzzy system is marked as a fuzzy system with fuzzy conclusions (the consequence of each rule is given by the statement).

In the second type of a fuzzy system there is a consequent rule presented as a combination of input variables. Rules are therefore given in the following way [3]:

$$\mathit{If}\ \left(x_1 = A_1^{jk}\right)\ \mathit{and} \ldots \mathit{and}\ \left(x_n = A_n^{jk}\right)\ \mathit{then}\ y = f_k(x_1, x_2, \ldots, x_n)\ \ k = 1, 2, \ldots, r \tag{3}$$

Plain linear combinations of input variables are quite often deployed as functions f_k.

$$f_k(x_1, x_2, \ldots, x_n) = a_{k0} + a_{k1}x_1 + a_{k2}x_{2_+} + \ldots + a_{kn}x_n \qquad k = 1, 2, \ldots, r \tag{4}$$

Then the overall output value is the weighted average of the particular fuzzy system rules outputs.

$$y = \frac{\sum_{k=1}^{r} w_k f_k(x_1, x_2, \ldots, x_n)}{\sum_{k=1}^{r} w_k}, \tag{5}$$

where weight w_k is the resulting value of the k-rule antecedent. [3]

1.2 Fuzzy System Selection

As in the environment of the socio-economic and qualitative intelligent transport systems impacts' estimation in majority of cases it is not possible to describe the consequence of implemented rules by a linear or any other combination of the input parameters, it is reasonable to use the Mamdani fuzzy system for these purposes.

2 Fuzzy/Simulation Model Deployment Framework

Basic principle of the fuzzy model deployment in the area of ITS impacts' evaluation lies in a separate set up of a fuzzy-linguistic model for each monitored parameter/magnitude. After further consideration it is necessary to expand this basic thought by the need of dedicated adjusting not only from the perspective of the particular estimated impact but also from the view of the ITS application. One of the steps leading towards the fuzzy modelling deployment is therefore the ITS architecture analysis which ensures the identification of the common representatives of different transport-telematics domains.

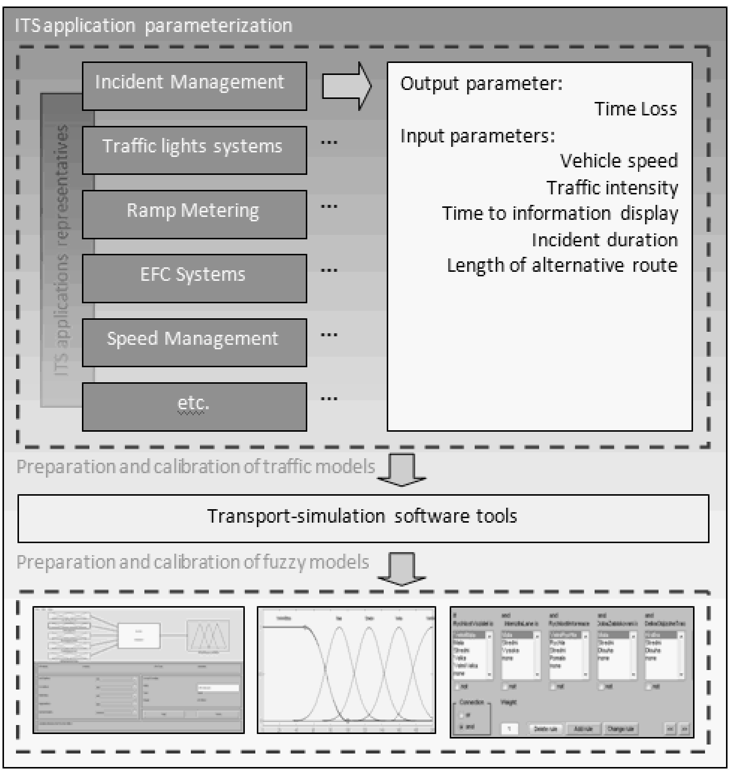


Fig. 1. Fuzzy/simulation model deployment framework

Each fuzzy model should be rigorously adjusted and calibrated. It has to be done not only from the perspective of the partial membership functions responding to the particular fuzzy sets but also from the view of inferential rules system definition. Again, it faces the requirements of relevant input data accessibility at this point. In this case inputs are not demanded for purposes of assessment itself but for the fuzzy model calibration and pre-set. A calibrated model will be capable to make an estimation of the wanted output value in a close relation to the inputs given in the form of fuzzy linguistic variables. Therefore, particular ITS applications have to be, for the purposes of the fuzzy model pre-set, on the side of inputs appropriately parameterized. It is important to undertake this parameterization taking into account the desired output parameter in the way that covers all key factors, which can influence its output values. It is also necessary in this context to take into consideration the rising complexity of a fuzzy model.

Values for the model calibration may be obtained on the basis of the expert knowledge, measurements of the already deployed systems and generalization of the observed values or by utilization of the transport simulation software tools. As mentioned in the introduction to this paper the deployment of the transport simulation software is under the main focus. The overall approach of the fuzzy/simulation model deployment framework is shown on the picture above.

2.1 Socio-Economic Evaluation - Example

In the subsequent paragraphs the model of transport-telematics application is introduced. It is used to present the procedures of the fuzzy model creation, which are intended for the evaluation of ITS benefits. For this purpose the Motorway Incident Management application is thought over. Its function consists in the identification of a traffic incident on the highways and motorways. After the interpretation of the transport-engineering parameters pointing out this situation and their validation the particular information is provided via variable message signs (VMS) to drivers. The VMSs provide the information describing the nature of incident and also give alternative routing. Key benefits are represented here in the form of time savings.

Simulation of an ITS application. The functionality of Incident Management ITS application was simulated in the micro-simulation software package PARAMICS. The model simulates an incident on a motorway, which after a few minutes blocks both lanes.

After certain time this incident is identified and related information together with the alternative routing provided to the drivers. The simulation model compares variants with and without the ITS deployment. A model representation is shown on the picture below.

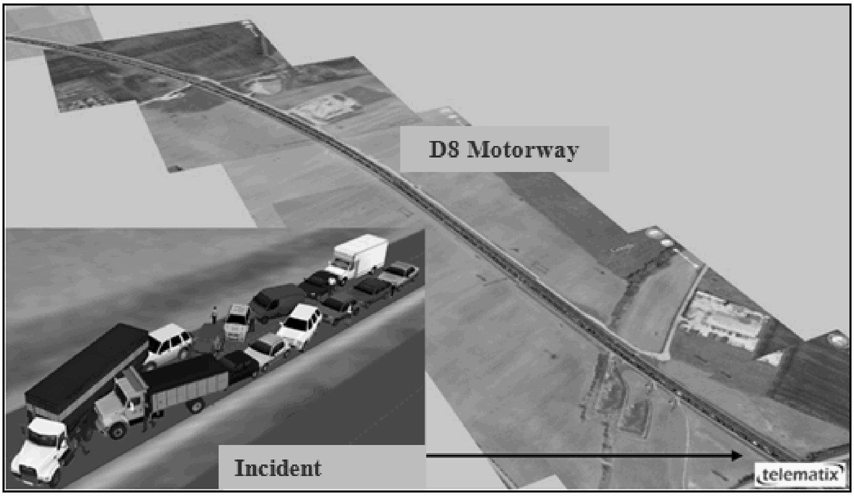


Fig. 2. Micro-simulation model – PARAMICS

Identification of basic parameters. With reference to the beginning of this chapter it is necessary to identify the input parameters, which will directly affect the output value of the time costs. In connection with the Incident Management functionality as described in 2.1 the following input parameters are taken into account:

- Vehicle Speed on the motorway: it directly influences the speed of traffic congestion occurrence.
- Traffic Intensity: second factor directly contributing to the speed of traffic congestion occurrence.
- Time Delay between incident and information provision to drivers: affects the length of congestion.
- Duration of motorway blockage affects the length of congestion.
- Driver Acceptance of provided information: User acceptance coheres with the value of so called driver-acceptance parameter which represents the percentage of drivers which reflect the information provided (via VMS) and affects the length of congestion.

In general, the requirement to describe two components influencing an ITS system as a whole is always valid. Firstly, the transport-engineering component and secondly the functional-context component needs to be defined. The transport-engineering component in this particular case is filled in by the Vehicle Speed on the motorway and the Traffic intensity. The second component – by the Time Delay between the incident and information provision to drivers;, the Duration of motorway blockage affects the length of congestion and the Driver Acceptance of information provided.

Fuzzy model creation process. The fuzzy-linguistic model was prepared in the software environment of the Matlab – Fuzzy ToolBox. The above defined parameters were taken as the basic inputs and per-partes transformed into the form of linguistic variables, terms' sets and universe adjustment.

The terms' sets definitions of the particular input parameters reflect the nature of linguistic variables and its common way of its lingual representation. It touches the consistency of fuzzy sets in this way, which is related to the number of defined terms. Therefore the number of defined terms varies from a fuzzy set to another one. This is also one of advantages the fuzzy-linguistic approximation brings to.

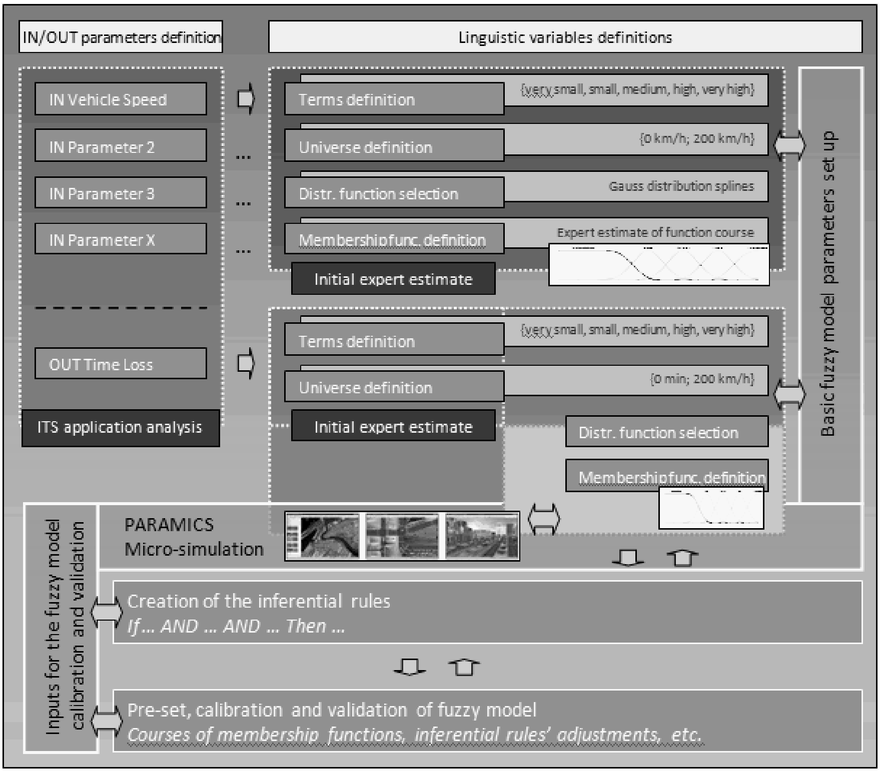


Fig. 3. Fuzzy model creation process

Basic pre-sets of particular membership functions were undertaken by the utilization of a Gauss distribution spline. This kind of distribution in these cases is rated as the suitable initial point for the fuzzy model creation. In the case of extreme terms' values (very small, very high) the distribution is modified into the form of continuous Z or S spline. It allows setting up the membership values in intervals to 1.00. Default adjustments of the partial membership functions were performed on the basis of expert estimation and common understanding of individual terms in the context of linguistic variables. Changes and modifications of the distribution functions' courses were performed as a part of validation.

On the output linguistic variable side the situation is more complicated. The initial pre-set of partial membership functions courses is not easily deducible. Therefore the PARAMICS micro-simulation model runs were processed.

From the figure above it is noticeable that after the basic adjustments of the input and output linguistic variables the definition of the inferential expert rules follows. These rules lead towards an n-dimensional fuzzy model on which background after the defuzzyfication the estimated values of the output parameter (time losses) are achieved.

Before the fuzzy model is ready to use it is necessary to undertake the calibration and also the validation procedures. It leads to the modifications of the partial membership functions types or courses and changes in the inferential rules mechanism. The overall process of the fuzzy model preparation is shown on the picture above.

Definition of input variable – Vehicle speed on a motorway. The definition of the input variables is shown on the example of the Vehicle Speed on a Motorway. All the other input linguistic variables were defined in a similar way.

The definition of this linguistic variable is based on the research activities of the Czech Road and Motorway Directorate published in 2006 [2]. These activities were focused among others on capturing the relation between the vehicles point speed and its occurrence frequency. This graph, as shown below, is the core source for the universe adjustment and the terms' set definition. It also supported the terms' representation in the form of partial membership functions.

Linguistic variable definition in the Matlab FuzzyToolBox is presented in Fig. 5.

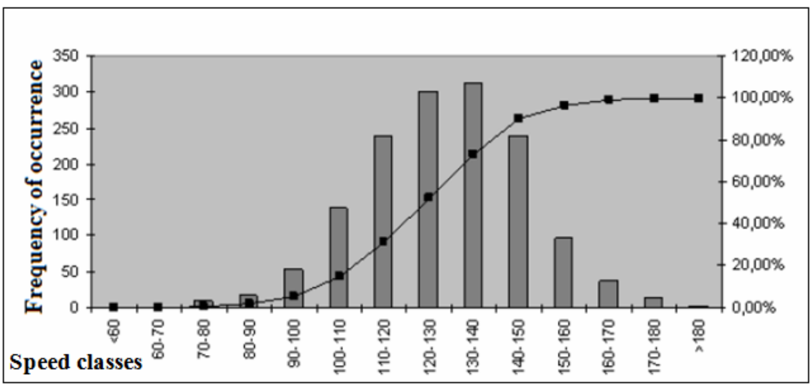


Fig. 4. Dependency of the vehicle frequency and its actual point speed on motorway D1 [2]

The definition of the linguistic variable is then as follows:

- Terms' set for the Vehicle Speed: {very small, small, medium, high, very high};
- Terms are defined on the universe of: <0 km/h; 200 km/h>;
- The partial membership functions courses are given as mentioned earlier by common Gauss distribution splines with maximums at 90, 130 and 170 km/h. Left and right ends of the universe are covered by Z or S spline with interval maximum 0-60 km/h and 200+ km/h.

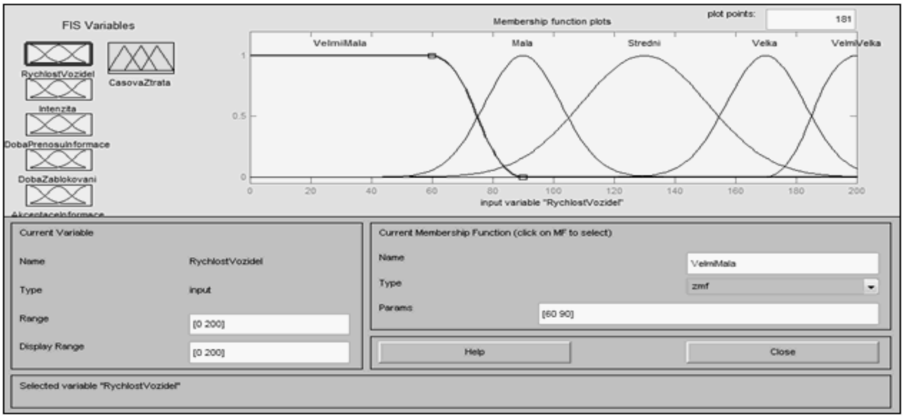


Fig. 5. Fuzzy linguistic variable definition – Vehicle Speed on Motorway

Definition of output variable – Time loss. The definition of output variable was, as stated above, given by the use of the software micro-simulation environment PARAMICS. The reason to do so is that the definition range of the variable (universe) as well as the terms' definitions cannot be provided expertly. All of these values have to be gained from the simulation model outputs.

The universe range was acquired on the basis of systematic experiments with the input parameters set up. For these purposes the extreme values were chosen (for both edges) and the universes' interval range was identified. The inputs to get the maximum value were adjusted as listed below:

- Vehicle speed: 130 km/h (small)
- Traffic intensity: 2000 veh/hour/lane (very high)
- Time delay: 40 minutes (very high)
- Time of motorway blockage: 65 minutes (high)
- User acceptance: 20% (very small)

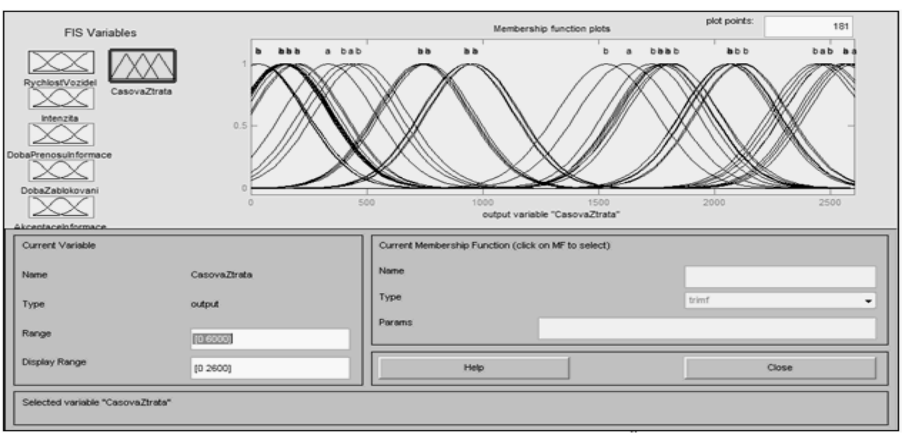


Fig. 6. Fuzzy linguistic variable definition – Time Loss

For this input values 10 runs of the simulation model with different seeds were performed. Outputs were averaged. After the data interpretation the Time Loss universe definition was given as <0 hours; 6000 hours>.

The inferential rules mechanism was prepared on the basis of the test scenarios matrix. This matrix is represented by reasonable combinations of the input parameters and the output value in the form of Time Loss is given by PARAMICS. The test matrix consists of 54 scenarios, which will give basic not only to inferential rules but also to the partial membership functions courses (see figure above).

Fuzzy model optimization and calibration. The fuzzy model optimization and calibration was performed reflecting the basic principles of fuzzy modelling. The number of partial membership functions was reduced and also the point of intersection was calibrated.

The amount of terms was reduced by elimination of all partial functions which extremes were closer than 15 hour/accident (maximums). It leads to the reduction of approximately 40% of partial membership functions and therefore to a better power of the fuzzy model. The number of inferential rules stayed the same to have a good coverage of universe.

The point of intersection for all the rest of the partial membership functions was set to the value of 0.5.

The final representation of the output variable is on the picture below.

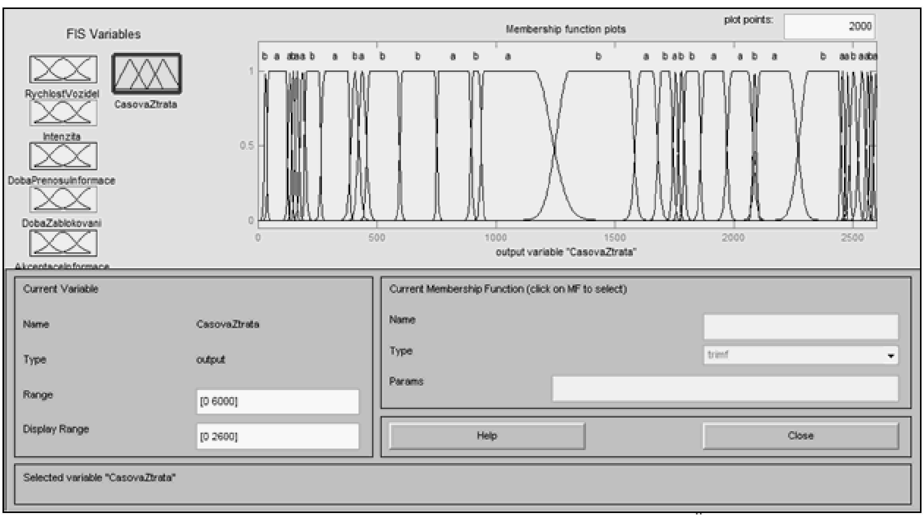


Fig. 7. Optimized fuzzy linguistic variable definition – Time Loss

3 Fuzzy Model Error Rate and Conclusion

Error rate of the proposed method for the evaluation of the socio-economic and qualitative impacts of the ITS application was enumerated by comparison of the micro-simulation software outputs (validation scenarios) and the outputs of the fuzzy model (given by the centroid method).

The overall error rate is within the interval <0%; 10%>, which from the perspective of ITS application evaluation is a result more than satisfactory. This rate can not affect and distort the overall values of cost-benefit analysis (CBA) parameters. Absolute values of the possible error in the fuzzy calculations are, as compared to the total Net Present Value (NPV), insignificant.

Therefore as the final conclusion it is possible to state that a brand new approach to the ITS socio-economic and qualitative benefits' evaluation was introduced in this paper. This method forms a valuable input to the CBA calculations and helps to cover more aspects related to the ITS application evaluation as a whole.

References

1. Svítek, M., et al.: Výzkum účinnosti telematických systémů v dopravě: Research report 2008. 01/2009 (2009)
2. Horníček, K., Dolejší, M.: Modelování dopravního proudu na vícepruhové komunikaci: Conference proceedings: Doprava, zdraví a životní prostředí. Centrum of Transport Research. 10/2006. s. 117–124 (2006) ISBN 80-86502-33-3
3. Jura, P.: Základy fuzzy logiky pro řízení a modelování, p. 132. VUTIUM publishing, Brno (2003)

The Conception Approach to the Traffic Control in Czech Cities – Examples from Prague

Tomáš Tichý[1] and Dušan Krajčír[2]

[1] ELTODO dopravní systémy s.r.o.,
Novodvorská 1010/14, 142 01 Praha 4, Czech republic
tichyt@eltodo.cz
[2] ELTODO EG a.s.,
Novodvorská 1010/14, 142 01 Praha 4, Czech republic
krajcird@eltodo.cz

Abstract. Modern and economic development of contemporary towns is without question highly dependent upon traffic infrastructure progress. Automobile transport intensity is dramatically rising in large towns and other Czech and European cities. At the same time number of traffic congestions and accidents is increasing, standing times are becoming longer and ecological stress is also escalated. To solve this situation seems to be the most effective solution to design intelligent traffic light intersection control system, variable message signs, preference of public transportation, road line traffic control and next telematics subsystems. This control system and subsystems should improve permeability of traffic road network with a respect for all demands on recent trends of traffic development in towns and regions.

Keywords: traffic, ITS, telematics, control.

1 Introduction

A traffic expansion in the cities is continuously rising, which fact could be seen in regular annual reports in the Czech Republic. For instance the traffic growth in Prague since 2000 is 36%. In the whole Czech Republic the traffic growth on highways and 1st, 2nd and 3rd class roads reached 35%. The growth of the cars per a citizen in the Czech Republic since 2000 is 16%. The number of cars in Prague is 1 vehicle per 1,8 citizen, which classifies Prague as a city with one of highest number of cars per citizen in Europe. This trend could be seen also in other cities in the Czech Republic but also in conjunction with economic development in other countries of Central and Eastern Europe.

One effect of this grow trend is the traffic intensity increasing in the cities and creating congestions. The effort to take the cars out of the city isn't always successful. Concerning a big traffic constructions such as circumferential highways, tunnel constructions and new road constructions and their initiation and termination dates the work period duration varies between a year and tens of years. The reason is not only a difficult engineering preparation but also a property buy out, proper permits ensuring, studies processing and not least gathering together the finance for the communication construction itself. If a high quality network of communications is not created it is

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 410–417, 2010.

necessary not only to plan but also to solve the sustainable current traffic network, which contains also own crossroads. In the light of traffic safety the rotary intersections could be projected on these communications. These rotary intersections have got some capacitive limitations. Another possibility to ensure the traffic safety is the installation of the traffic lights whose capability is much better than in the case of rotary intersection and whose safety is distinctively higher than in joining intersections.

Light-controlled intersections enable some operative functions, which have important effective impact upon the traffic such as redirecting, impeding, preferences, demanding of chosen direction etc. More effective controlling significantly increases the intersection´s capability and the transmissivity is raised c. by 20%. If several intersections with specific distance no longer than 750m – 1000m are optimally suggested, it is appropriate to integrate these intersections into one coordinated package in order to make the transit during the area more effective. A network transmissivity increment to 10-20% could be achieved by using above described procedures according to the assigned area or communication.

2 The Traffic Control in Prague

The capital city of Prague was historically divided into 10 basic boroughs, which were further divided up to 22 boroughs. These boroughs are in the traffic domain still evolving. Traffic lights addressing is resulting from the older distribution of the boroughs i.e. 10 boroughs.

Fig. 1. Schematic layout of TCC in area in Prague with new area marking

Gradually a demand for current number of traffic light controlled crossroads expanding is rising. Also a stepwise integrating of traffic light controlled crossroads into current zone centers and requirements to build new regional traffic store controllers connected into a central junction of traffic controlling are required. A demand to define individual areas especially such kind of marking able to clear the sub region and being naturally rational is offered (Fig. 1).

Changing the area marking from numeric to alphabetic means not only definite area marking with regard to the boroughs but also logic locating in the city area. A change of the area marking is presented in Tab. 1.

Table 1. New area marking

Old marking	City areas	New marking	Acronym
Area 1	Holesovice	Central 3	C3
Area 2	Dejvice, Ruzyne	North-west	SZ
Area 3	Smichov, Brevnov	Central 2	C2
Area 4	Stodulky, Repy	South-west	JZ
Area 5	Stare Mesto, Karlin, Zizkov	Central 1	C1
Area 6	Podoli, Pankrac, Modrany	South	J
Area 7	Strasnice, Malesice	South-east	JV
Area 8	Vysocany, Bechovice, Hloubetin	East	V
Area 9	Prosek, Kobylisy, Bohnice	North	S

Area borderlines are defined on the basis of a certain historical solidarity whereas integration of some crossroads especially the ones in the area border or in its immediate environment is going to be judged considering traffic and technical possibilities of the situation.

Decisive condition to manage successfully the traffic congestions with protecting traffic safety is to control a system able to enable the traffic engineers to use effective automatic driving instruments with the highest elasticity level. Simultaneously this system has to be easily controllable and reliable, allowing the operator to concentrate on optimizing the driving time, increasing the quality of transmissivity through the area, and reacting to incurred events and as much as possible effective improving of the traffic and transport process control. The hearth of the traffic lights central controlling system is a SITRAFFIC SYSTEM. This distributed controlling system between traffic control units and central controlling system provides excellent reserve structure in case of any communication problems between these two components. This kind of distribution is also useful for everyday traffic control. An operating activated by local (microscopic) traffic combined with centralized (macroscopic) traffic driving gradually gains control of central monitoring and traffic light controlled crossroads driving in the capital city of Prague. All the current control units will be managed and coordinated from the central dispatching. Primary functions are visualization, measured values processing, archiving, report and driving, which are the base of the traffic central office. SITRAFFIC SCALA system works with GIS application on which all management including summary of events etc. could be done (Fig. 2).

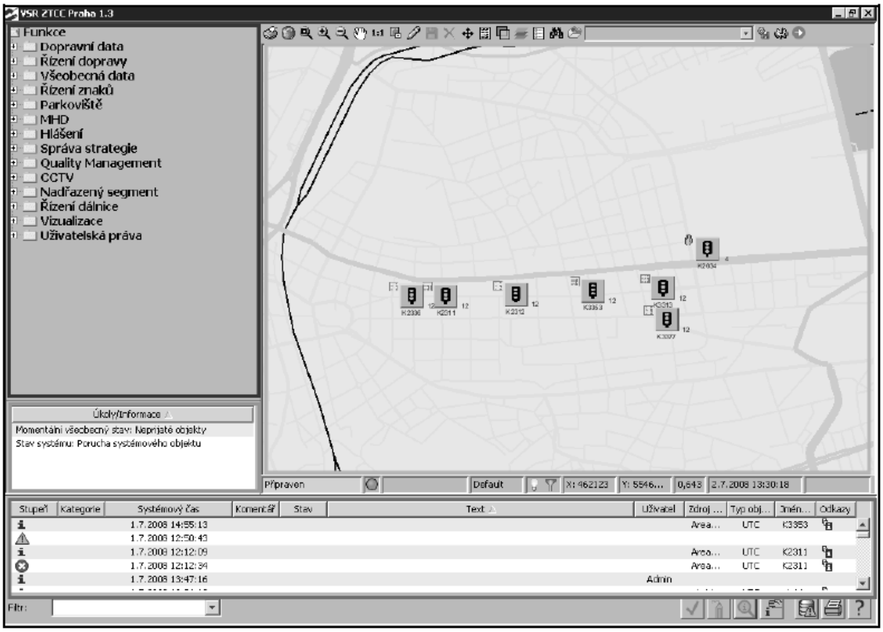


Fig. 2. A GIS application from the SITRAFFIC SCALA system

3 The Reliability of Control System

To secure the system function and whole control system reliability is necessary to set criterions for maintenance, exchange and new technology delivery. Replacing the old technology means increasing of the reliability and also reducing the operative intervention necessity. It is related especially to poles, cables but also equipment such as light bulbs, detectors etc. exchange. The equipment failures are caused by the age of the components, random accidents hard to predict such as weather or failures of electronic components themselves. Separate part is also harmful events such as vandalism, weather influence and accidents on the communications. While changing the old equipment lot of damages caused by the traffic participants or weather influents could happen. These damages significantly exceed the amount of the common failures and are also more time consuming.

3.1 The Failure of Equipment

A long term objective is to replace technologically old and defective equipment and to increase the reliability of the whole system. It means especially a new control units and LED technology signal device installation, using new detectors but also cables and poles. On the Fig. 3 an evident growth of the harmful events replacing time could be seen, especially in may 2009 which was caused by climatic influences. The replacing time is also protracted by big accidents whereat a pole and especially control unit and cables are damaged. On the other hand the smaller range failures are repaired on average in 30 minutes.

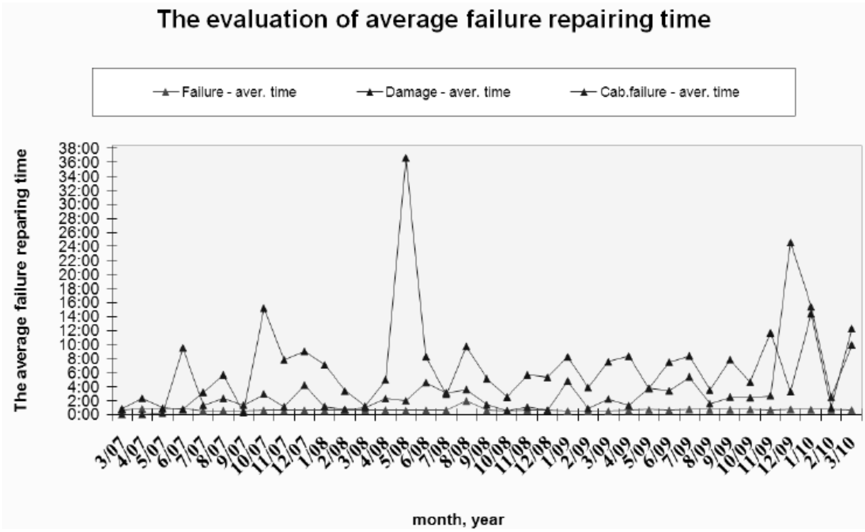


Fig. 3. Average failure repairing time

An example of failure week distribution is demonstrated on the Fig. 4. A similarity on failure week distribution and course of traffic situation in Prague especially in the morning could be seen on the figure.

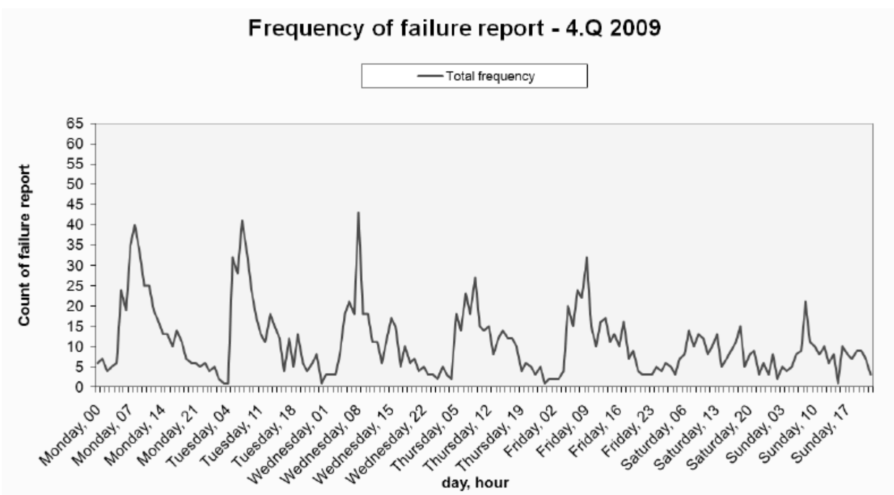


Fig. 4. Failure week frequency

Most frequent failure are the broken light bulbs (up to 54%), which are gradually replaced by LED technology. Other examples are detector failures (c. 15%) and outdoor equipment failures caused also by harmful events which significantly boost the equipment failures number and also already mentioned failure effect replaces time extension.

3.2 Emergency Services

Average arrival time to the failure place is steady about 20 minutes. While extraordinary situations a small time increase could be achieved. This delay is caused by outdoor influents such as strong wind, rainstorm etc. A delay could be achieved also because of difficult traffic situation i.e. no access to the incident place. Since the priority is to remove the more important or more extensive accidents, the maximal arrival time is achieved in the case of failures without any influence to the device function.

The Fig. 5 shows quick arrival to the failure place where the arrival time limit was set 1 hour from the failure report but the failure repairing was quick. A replacing time distribution is described in Fig.3.

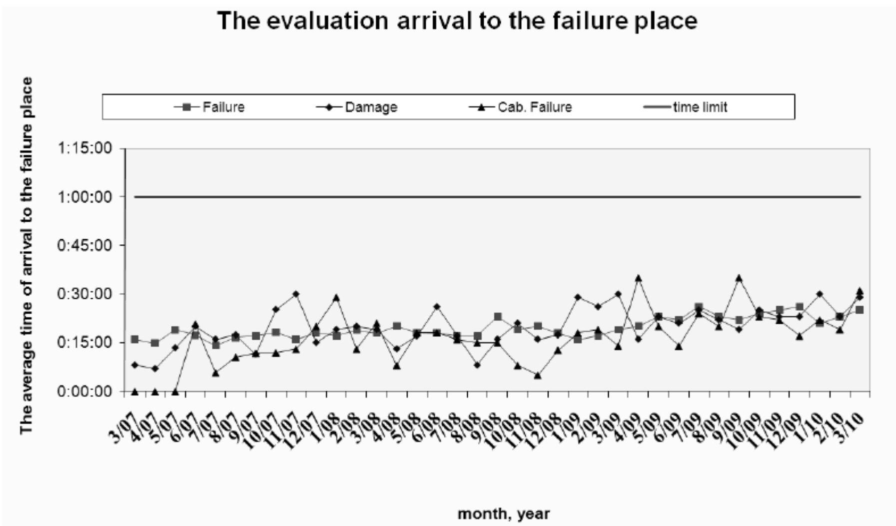


Fig. 5. Operative unit arrival to the failure place

4 The Traffic System for Control in a Town

A part of the traffic control central - SITRAFFIC SCALA is a monitor of state equipment and control part too, enable to intervene to the traffic control in the area of the city. The system has got a control algorithm over the manage area. The objective is to apply the higher control algorithms for network application of automatic traffic control in the city area by using the traffic lights in intersection points using MOTION and TASS systems [6]. The other control system suitable for recommendation is the module for strategic control in city – STRAMO. This system is instrumental to manage operatively the single areas interacted each other.

Module STRAMO – Strategy module ensures a strategic control at the city level using logic conditions. The system integrates heterogeneous data from detectors, information and from other sources of telematics systems. The data are saved and compared with states of equipments by the module of Situation analysis. The output of the analysis state is an individual situation which is simultaneously the input to the module of logic control. In this module of logic control the parameters of strategy control could be modified and edit. The operator has got the same competence. The module of action sends the scenario to the selected telematic equipments. The scenario is sent to the archive for next using.

Module STRAMO does not substitute the TASS or MOTION system, but then it supports and extends as an complex traffic control of the intersection points. The module allows to selects a strategy in frame of area control, intersection points, main routes, etc. One can use for special situations in traffic like congestions, incidents in traffic net, accidents etc.

The new methods of traffic control in area are on the basis of length of queue estimation or optimization of the transit time through the area. The suggestion of control model could be put to test in appropriate simulation tool. The Scala system has open an interface for cooperation with new models of control and systems (VMS, cameras, parking etc.). In the future it would be necessary to use the data from a floating vehicles (BUS, TAXI etc.) to get the data for traffic control model and information about problems in traffic net as well.

5 Conclusions

The suitable solution for a traffic control in the cities is a technology or more likely a combination of technologies able to increase a possibility trough a city network preferably with a low environmental stress and able to react very quickly while a fluent and safe traffic. The aim of this article was to describe issues of traffic control in urban areas and to offer possibilities how to solve this problem.

It is necessary for global traffic control system in town to connect telematics system to the traffic centre. The centre has to have a quick failure diagnostic and has to enable an immediate response to the incident. The system is able to diagnose the most responsive element or elements of the telematics systems in the traffic network and to make an operative intervention in time. Very important is the periodic prevention, service and changing of the technology. If the higher algorithm of the traffic control for example STRAMO is used the automatically connection of the telematics systems and the control in a traffic network in the city could be improved.

In the future will be necessary to design telematics system using the new technology, like data collection by floating cars etc., however a service method of this system will be necessary to create.

References

1. Mueck, J., Hanitzsch, A., Condie, H., Bielefeld C.: Signal management in real time for urban traffic networks. SMART NETS, Germany (2004) IST-200-28090
2. Hounsell, N.B., Shrestha, B.P., Piao, J., McDonald, M.: Review of urban traffic management and the impacts of new vehicle technologies. IET Intell. Transp. Syst. 3, 419 (2009)
3. Přibyl, P., Svítek, M.: Inteligentní dopravní systémy, BEN, Praha (2002)

4. Povejšil, L.: Pravidelná hodnotící zpráva o dosahovaných výsledcích na zakázce Praha, Dokumentace ELTODO, Praha (2010)
5. Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., Wang, Y.: Review of road traffic control strategies. Proceedings of the IEEE 91(12), 2043–2067 (2003)
6. Tichý, T.: Vyhodnocení funkčnosti systémů řízení dopravy v oblasti města. Silniční obzor, 5/2007, ročník 68, číslo 5, Česká silniční společnost, pp. 132–137
7. Tichý, T. Povejšil, L.: Systém regulace městské dopravy v Praze. In: ITS Bratislava 2008, WIRELESSCOM s.r.o., Bratislava (2008)

Algorithm of Unmanned Aircraft Systems Displacement in Airspace*

Tomasz Gugała

Combined Air Operations Center Uedem, NATO,
Mühlenstr. 89, 47589 Uedem, Germany
pink_dragon1@o2.pl

Abstract. Despite the fact Unmanned Aerial Vehicles have been used for more than 70 years and their uncommon development has taken place in the first decade of the 21st Century, there is still no elaboration of „Uniform Concept of the Unmanned Aircraft Systems Displacement in Airspace”. The indispensable condition of the above mentioned concept has to be flight safety of all airspace users. To achieve this goal, it is necessary to work out the adequate procedures and regulations in the scope of airspace usage taking into consideration this up-to-date means of air transport. Therefore, elaboration of the algorithm by the author, can be a reason of achievement for the above mentioned object in the near future. Under such circumstances, the author has taken the trial to perform this challenging task.

Keywords: Aircraft, Unmanned Vehicle, Algorithm Systems.

1 Introduction

Based on the analysis of the available literature of this subject, participation in international conferences and seminars, as well as personal experiences of the author achieved during many years of service within the NATO command structure, it is allowed to impose the thesis, that UASs are currently one of most dynamically developed contemporary armament systems, not only within the North Atlantic Treaty Organization, but also in many countries out of the Pact. From year to year the new concepts of UAS adaptation are generated in a wide spectrum of military services. It is also undertaking the task of transforming the new modern technological achievements in this domain to the civilian environment. The above mentioned concepts of military applications are generating many miniaturized, and the large UAS constructions which are able to realize the following tasks:

---

* *UAS* – abbreviation is recognized as one of many related terms (*UAV* – Unmanned Aerial Vehicle; *UAVS* – Unmanned Aerial Vehicle System; *UCAV* – Unmanned Combat Aerial Vehicle; *RPV* – Remotely Piloted Vehicle; etc).

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 418–426, 2010.

- Intelligence, Surveillance and Reconnaissance – ISR;
  a) Stand off,
  b) Over flight,
  c) Denied Access.
- Suppression of Enemy Air Defense - SEAD;
- Air Attack (Strike);
- Destruction of Enemy Air Defense – DEAD;
- Electronic Warfare Attack;
- Anti – Surface Warfare;
  a) Maritime Interdiction,
  b) Counter Piracy.
- Underwater Warfare;
  a) Anti Submarine,
  b) Counter Mine.
- Communications Relay.

In order to effective utilize all available UAS constructions[1], NATO experts have accomplished the new classification of UAS on respective classes and categories in dependence of weight, way of employment, operating altitude, mission radiusand primary supported commander (Table 1.). Still in the year 2008, partition of NATO UAS has been involved three categories (TACTICAL, MALE, HALE) and current obligatory partition is involving three additional categories (SMALL, MINI, MICRO) as well as three classes, where the additional criterion of partition is the weight (>150kg; 150 – 600kg; <600kg) (Fig. 1.).

Commonly accepted and adequate defined UAS categories establish the foundation for NATO UAS terminology. These categories facilitate communication and knowledge by delivering unifying standards for different kinds of organizations and institutions. Suitable NATO organizations refine standards and doctrine which increasingly include UAS considerations. Those considerations into NATO STANAGs[2] are better served with a common reference system. The categories can improve NATO operational planning and C2 (Command and Control) by providing a common reference for grouping UAS. During realization of the tasks in Non–segregated areas, UAS classification can be helpful in case of flight safety criteria elaboration, categorization of the respective groups of UAVs and certification of UAS personnel by analogy to air crew personnel certification and categorization of manned aircraft. Therefore, STANAGs are categorizing all UAS family, according to common defined criterions in the aim of the most effective application.

Based on this every country being in possession of a UAS fleet should organize training and equipment of those forces as well as the rules of combat employment.

---

[1] More than 60 operational models of Unmanned Aircraft, more than 2200 ground control segments, over 6700 operational Unmanned Aircraft in NATO.

[2] NATO Standardization Agreement (STANAG – 4586) for UAVs is a basic document continuously updated.

Table 1. NATO UAS Classification Guide. September 2009 JCGUAV meeting.

UAV CLASSIFICATION TABLE						
Class	*Category*	*Normal employ-yment*	*Normal Operating Altitude*	*Normal Mission Radius*	*Primary Supported Comd.*	*Example platform*
CLASS I (less than 150 kg)	*SMALL* >20 kg	Tactical Unit (employs launch system)	Up to 5K ft AGL	50 km (LOS)	BN/Regt, BG	Luna, Hermes 90
	MINI 2-20 kg	Tactical Sub-unit (manual Launch)	Up to 3K ft AGL	25 km (LOS)	Coy/Sqn	Scan Eagle, Skylark, Raven, DH3, Aladin, Strix
	MICRO <2 kg	Tactical PI, Sect, Individual (single operator)	Up to 200 ft AGL	5 km (LOS)	PI, Sect	Black Widow
CLASS II (150 kg to 600 kg)	*TACTICAL*	Tactical Formation	Up to 10,000 ft AGL	200 km (LOS)	Bde Comd	Sperwer, Iview 250, Hermes 450, Aerostar, Ranger
CLASS III (more than 600 kg)	Strike/ Combat	Strategic /National	Up to 65,000 ft	Unlimited (BLOS)	Theatre COM	
	HALE	Strategic /National	Up to 65,000 ft	Unlimited (BLOS)	Theatre COM	Global Hawk
	MALE	Operational/Theatre	Up to 45,000 ft MSL	Unlimited (BLOS)	JTF COM	Predators A and B, Heron, Heron TP, Hermes 900

2 Main Topic

Having mentioned the above information and UAS classification within the military structures let us apply the similar criteria and philosophy of the UAS adaptation in the civilian environment. The first trials and ideas of UAS employment out of military service are relating to the following groups of tasks:

- Commercial Cargo Transport - CCT;
- Humanitarian Cargo Transport;
- Search and Rescue – SAR (Air, See, Land environment);
- Society Protection against the effects of Natural Disasters;
  a) Flood, Drought,
  b) Earthquake,
  c) Avalanches,
  d) Fires,
  e) Hurricanes,
  f) Volcano Eruption.
- Airspace and Border-crossing Patrolling;
  a) Refugees,
  b) Drugs Smuggling.
- Environmental Protection:
  a) Tanker Leaks,
  b) Industrial Waste,
  c) Contamination Areas,
  d) Deforestation.
- Climate Changes Observation;
- Weather Monitoring;
- Property Protection;
- Police "Eyes in the Sky".

Nevertheless, realization of the mentioned tasks above is continually extremely difficult because there is still no existing "Uniform Concept of the Unmanned Aircraft Systems (UAS) Displacement in Airspace". The main goal of the related elaboration should be flight safety assurance of all air vehicles as well as in national and international airspace. This issue is highly complicated. Therefore, despite a number of investigations, initiatives and advanced projects of many international experts, there is still no desired final effect. Today, there are no less than eight major NATO, civilian organizations and working groups who are addressing various aspects of integrating UAS into Non-segregated area in and amongst the manned fleet. Currently, the above mentioned organizations[3] are also working on many aspects heading to UAS integration to military and civilian Air Traffic Management (ATM). The primary issue of their effort is the common application of partition of the airspace (decreasing of the airspace classes from

[3] NATO FINAS (Flight In Non-segregated Air Space) Working Group; EUROCONTROL; USA FAA (Federal Aviation Administration); NSA (NATO Standardization Agency); NATO JAPCC (Joint Air Power Competence Centre) Project Team on UAS; NC3A (NATO Consultation, Command and Control Agency); CNAD (Conference of National Armaments Directors JCGUAV (Joint Capability Group on Unmanned Aerial Vehicles; ICAO (International Civil Aviation Organization).

seven to three by the end of the year 2010), adequate ATM procedures, proper Command and Control Rules, Emergency Procedures, Weather Minima, Border Crossing Procedures and many other indispensable prescriptions.

The main topic of this paper is the authorship presentation of the „Foreseeing Algorithm of the Civilian Cargo UAS Displacement in Non-segregated Airspace" based on the current achievements in this domain (Fig.2.).

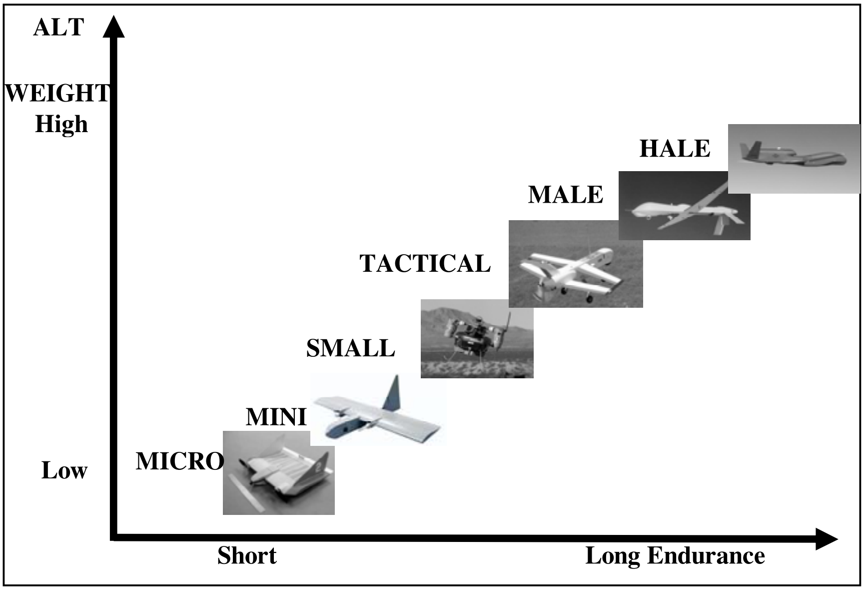


Fig. 1. NATO classification of UAS (Source: The JAPCC Strategic Concept of Employment for UAS in NATO 2010)

Consider this example: the UAS flight is proceeding within controlled airspace in class G and C, FL100, under control of civil GAT (General Air Traffic) from airfield ALPHA to airfield BRAVO. The weather conditions are as follows: no clouds, Visibility – 20 km. It is also assumed, these are cyclic over-flights (five times per day) of an Airlift Bridge, with a few tons load of food articles on-board. The distance between airfield ALPHA and BRAVO is about 1000 km and the time of one way flight is about two hours, including take off and landing procedures. The over-flight is within the airspace of one European country, and has been divided into seven phases:

Phase – 1. First, the UAS performs technical service before the flight [entering designated IFF (Identification Friend or Foe) code, checking of on-board devices, communications systems and remote control, refueling, anti-collision systems etc.].

According to approved flight plan [sent to the proper civilian ATM cell by the PIC-I (Pilot in Command)], is UAS engine „start up", taxi and take off, under control of airfield ALPHA TWR (Tower). The TWR provides two-sided radio contact (two independent radio channels of communication), and direct telephone line with PIC-I, who is situated in one of detached office of the Airport ALPHA. The PIC-I has also two independent direct telephone lines with the GAT controller.

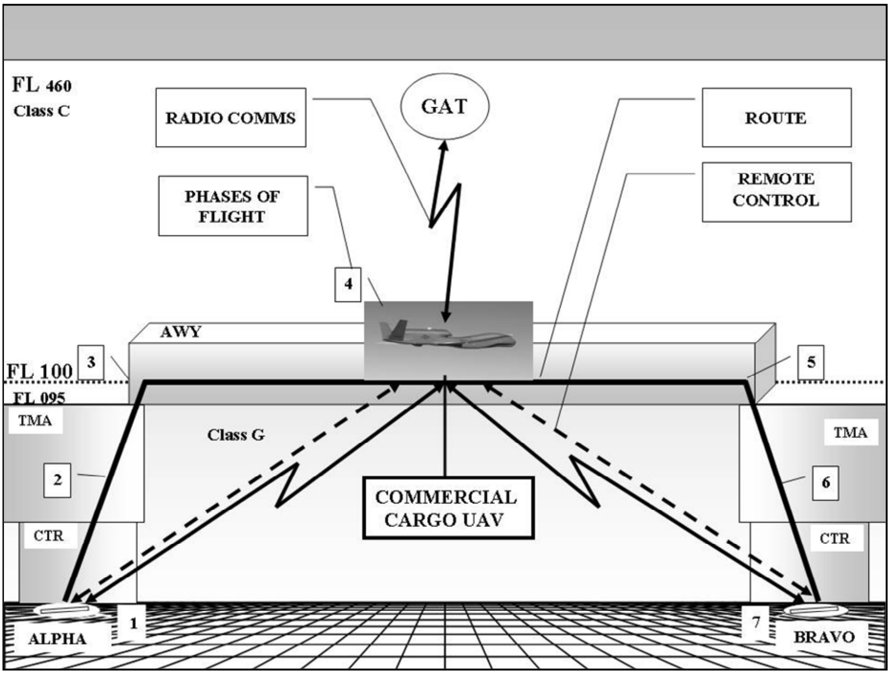


Fig. 2. Algorithm of the Civilian Cargo UAS Displacement in Non-segregated Airspace

Phase – 2. After take off, the ALPHA TWR controller passes by telephone the exact time of UAS take off, C/S (Call Sign) and IFF code (unchanging during the timeframe of the flight) to the GAT controller. The UAS climbs to FL100 in accordance with preprogramming flight conditions by the PIC-I. Before the UAS enters the AWY (Airway), PIC-I establishes radio contact with the GAT controller using the required frequency, declaring the entrance to the AWY with simultaneous remote control of all flight parameters.

Phase – 3. During the UAS flight within the AWY, PIC-I maintains the specific flight conditions [speed – 310 kt (TAS – Thru Air Speed); FL100; heading – 300º]. Simultaneously he keeps continuous two-way radio contact with the GAT controller by two independent radio stations and retranslator installed on-board of UAS. This solution enables uninterrupted radio communication between PIC-I and GAT, and remote control of the UAS.

Phase – 4. The UAS continues autonomous stabilized flight approaching airfield BRAVO at FL100. The PIC-I controls all flight parameters, and adepts if the GAT controller orders a change of UAS heading or altitude, as the operational air situation dictates. In case of any damage, emergency situation or loss of remote **control authority, the UAS** autonomously continues flight on the set the airfield BRAVO, evading every possible obstacle and then is diverting to the preplanned flight path, or breaks the air mission, flying to the alternate airfield or returning to home base, depending on it's autonomous emergency procedures.

Phase – 5. The UAS approaches the exit point of the AWY. At the same time the remote control is taken over by PIC-II out of airfield BRAVO, who is responsible for immediately establishing radio contact with the GAT controller. The GAT controller assigns the task to establish radio contact with airfield BRAVO TWR and to continue descending within TMA (Terminal Area) of BRAVO airfield.

Phase – 6. Aircraft autonomously continues descending to the airfield BRAVO, applying adequate landing procedures, and lands on the RWY (Runway). BRAVO TWR controller is responsible for ensuring the UAS has landing priority with the exception of Passenger Manned Airliners, always priority one.

Phase – 7. The UAS safety lands on airfield BRAVO, skillfully leaving the RWY, taxiing and stopping at the designated place. Engine shutdown follows, and all on-board devices and cargo are unloaded. The UAS is then refueled and prepared for the next air mission to return to home base, using the same rules and procedures.

The presumed equipment and technical data performance of UAS

Equipment:

- board retlanslator;
- equipment of remote and autonomous control [Flight control, Navigation and Redundancy, Structural Damage and Fault Tolerance, Auto Take Off and Landing System, Anti Collision Systems – Sense & Avoid (MIDCAS[4], TCAS[5]);
- Surveillance means (IFF, ADS-B[6]);
- TV camera.

Technical Data Performance

This is commercial Cargo UAS, type – “Global Hawk RQ-4” with the following data performance:

- *Power-plant:* Rolls-Royce AE3007H Turbofan. lb;
- *Dimensions:* Length 14.5 m, Height 4.6 m, Wingspan 39.9 m;
- *Weight:* MTOW 14.628 kg, Max payload 1360 kg;
- *Performance:* endurance speed 310 (TAS) kt, speed endurance 36 hr, ceiling 60,000 ft;
- *Payload:* Raytheon-Synthetic Aperture Radar, Electro-Optical Infrared;
- *Data-link:* Ku SATCOM, CDL, LOS,INMARSANT;
- Lau*nch & Recovery:* Wheeled;
- *Electrical Power:* 25 KVA;
- *Ground Control Station:* Launch & Recovery Element, Mission Control Element.

[4] MIDCAS – MID-air Collision Avoidance System. Its mission is to demonstrate the baseline of solutions for the unmanned aircraft system MID-air collision avoidance function being acceptable as well as by the manned and UAS aviation foreseeable to operate in Non-Segregated airspace by 2015.

[5] TCAS – existing Tactical Collision Avoidance System.

[6] ADS-B – Automated Dependent Surveillance – Broadcasts, radiates signals containing an AC type, identification, GPS position, altitude, heading, speed, intent (i.e. climbing, descending or FL) and other data.

On the end of this topic the author presents argue of the following solutions:

- Board Retlanslator application provides continuous radio relay between PIC-I and PIC-II on the ALFA and BRAVO airfields, as well as with the GAT controller without the need of SATCOM;
- Continuation of UAS flight on FL100 minimizes de-confliction with Passenger Airliners, which are fly mainly from FL150 and above;
- Application of TV camera in the front of the UAS lets PICs and the GAT- TWR controllers observe the airspace in front of the aircraft for safety reasons;
- Use of two or more Ground Control Stations of the UAS (standard), seems helpful to maintain the required flight safety conditions.

3 Conclusions

This article is only the opinion of the author for exchange of views. Therefore, this application is narrowed to a UAS flight in Non-segregated airspace, applying overall baseline information related to this issue (Commercial Cargo Flight). It seems to be also an original solution to utilize conventional means of communication instead of using SATCOM (providing there are no "line of sight" barriers).. Also the proposal of application of advanced automated technology system ADS-B which may replace radars in the near future, is an idea known most probably only in USA, Russia and Israel. It is the author's belief that in the near future routine, operations of civilian and military UAS in General Air Traffic will become a reality.

The ESA (European Space Agency) expects to prepare an Initial Operational Capability in the C2/ATC area for Europe by 2015.

References

1. Bernier, R., Filbert, P., Steed, D.: Standardizing Unmanned Aircraft Motion Imagery and the Implications for NATO. The Journal of the JAPCC edn.11 (2010)
2. Bickley, J. R.: UAS Integration into the European National Air Space System, JAPCC Flyer edn. 1, March (2010).
3. Borke, B.W.: Global Dynamic Operations – Allocation of Remotely Piloted Aircraft among Combatant Commands. Air & Space Power Journal, Spring (2010)
4. Brzezina, J.: Szkolenie Operatorów Bezzałogowych Statków Powietrznych, Przegląd Sił Powietrznych (February 2009)
5. Falkow, E.: UAS in Civil Airspace – Flights Under Existing ICAO Arrangements. Global ATM Forum on Civil/Military Cooperation, Montreal (October 2009)
6. NATO Standardization Agreement (STANAG 4586) for Unmanned Aerial Vehicles
7. The JAPCC Flight Plan for Unmanned Aircraft Systems in NATO (2008)
8. The JAPCC Strategic Concept of Employment for Unmanned Aircraft Systems in NATO (2010)
9. Unmanned Aircraft Systems Roadmap 2005 – 2030, Office of the Secretary of Defense, Washington D.C. (2005)

10. EUROCONTROL DOK,
`http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/trans/106966.pdf`
11. Flight Control, Navigation and Redundancy,
`http://www.rockwellcollins.com/athena/demos/shadow/index.asp`
12. Structural Damage/Fault Tolerance,
`http://www.rockwellcollins.com/news/page11697.html`

The Reliability and Safety of Railway Control Systems Based on New Information Technologies

Andrzej Lewiński and Tomasz Perzyński

Kazimierz Pułaski Technical University of Radom,
Faculty of Transport and Electrical Engineering,
26-600 Radom, Malczewskiego 29, Poland
{a.lewinski,t.perzynski}@pr.radom.pl

Abstract. The paper contains the safety analysis of the new generation of railway control systems designed with respect to new technologies: microcomputers connected via cable and wireless transmission networks. In the paper the *THR* and global safety criteria are introduced to designed, tested and exploited railway control systems. For one of the systems the verification of safety parameters was possible in the form of both forecasting estimation and exploitation result. The authors presented in the paper one of the methods of forecasting estimation of control and management railway systems.

Keywords: Safety of railway control systems, THR, reliability evaluation.

1 Introduction

Development of the computers technique as well as its implementation in railway cause the appearance of many solutions based on microcomputers. The implementation of new techniques solutions allows computer systems to be better diagnosed, and also makes the registration of failure or facilitation in service of devices. According to the PLK Report (*Polish Railway Line 2008'*) the computer railway and management systems were introduced and installed in 68 railway control districts, and cross level protection systems were equipped in 859 modern control computer systems. The old prior management and control systems are being replaced by modern computers, often the self - actively control systems. Taking under consideration the safety, such computer solutions must possess an exploitation certificate. Computer devices, which are working in heavy, differential conditions, have to possess at least the same level of safety like the previous systems.

The problem is that in new solutions of computer management and control systems the safety parameters are calculated on the basis of coefficient estimation. As a result of such estimation, the received values are only approximated. The verification of estimation is possible after a few years of systems' working. The coefficient analysis bases on special railway standards. They give information about permitted and forbidden electronic circuits and components as well as necessity of systems' redundancy.

In the paper there are both the safety analysis of cross level protection system *RASP-4* and new conception of such system in which wireless communication between components is used. In the paper the estimation of failure rate was calculated

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 427–433, 2010.

on the basis of *Military Hand Book* [6]. For both described systems the *THR* (*Tolerable Hazard Rate)* was estimated. The *THR* rate presents (1):

$$THR = \prod_{i=1}^{n} \frac{\lambda_i}{t_{d_i}^{-1}} \cdot \sum_{i=1}^{n} t_{d_i}^{-1} \tag{1}$$

where: λ_i – failure rate, $t_{d_i}^{-1}$ – safe down rate.

For the system, in which the testing is holding periodically, the safe down rate equals:

$$t_d = \frac{T}{2} + NT \tag{2}$$

where: T – time of periodical testing, NT – negation time.

Systems described in the paper belong to the group of without repair systems. In case of failure of components or card, the system goes to safety state. In such a situation the management and control are led by special railway rules. System is fully functional after repair and restart.

2 Forecasting Estimation as a Basic Method of Reliability Analysis

Forecasting estimation is one of the elements of system's safety proof. It is the basis of evaluation of safety parameters, especially before coming of the new systems in practice. In forecasting estimation the calculation is based on producers' modules configurations. In the case when the characteristics of reliability of elements are unknown, but the structure of module and what it is made of (electronic elements, microprocessors or type of scale of integration the electronic circuits) is known, it is possible to estimate the failure rate λ [1], [4]. On (3) the general forecasting estimation is presented [6]:

$$\lambda_p = \lambda_b \left(\pi_T \cdot \pi_A \cdot \pi_R \cdot \pi_S \cdot \pi_C \cdot \pi_Q \cdot \pi_E \right) \tag{3}$$

where:

- λ_p - the part failure rate,
- λ_p - the base failure rate usually expressed by a model relating the influence of electrical and temperature stresses on the part,
- π_E - environmental factor,
- π_A - application factor,
- π_S - electrical stress factor,
- π_T - temperature factor,
- π_R - power rating factor,

- π_Q - quality factor,
- π_C - contact construction factor.

With reference to other elements, there can be another π factors in (3), like in the examples below [6]:

- capacitor - $\lambda_p = \lambda_b(\pi_Q \cdot \pi_E)$,
- resistor - $\lambda_p = \lambda_b(\pi_R \cdot \pi_Q \cdot \pi_E)$,
- transistor – $\lambda_p = \lambda_b(\pi_T \cdot \pi_A \cdot \pi_R \cdot \pi_S \cdot \pi_Q \cdot \pi_E)$,
- LED - $\lambda_p = \lambda_b(\pi_T \cdot \pi_Q \cdot \pi_E)$.

2.1 RASP-4 System

For the *RASP-4* system (cross level protection systems produced by Kombud S.A. from Radom) safety estimation has been done on the basis of exploitation results. The basic structure is presented in Fig. 1.

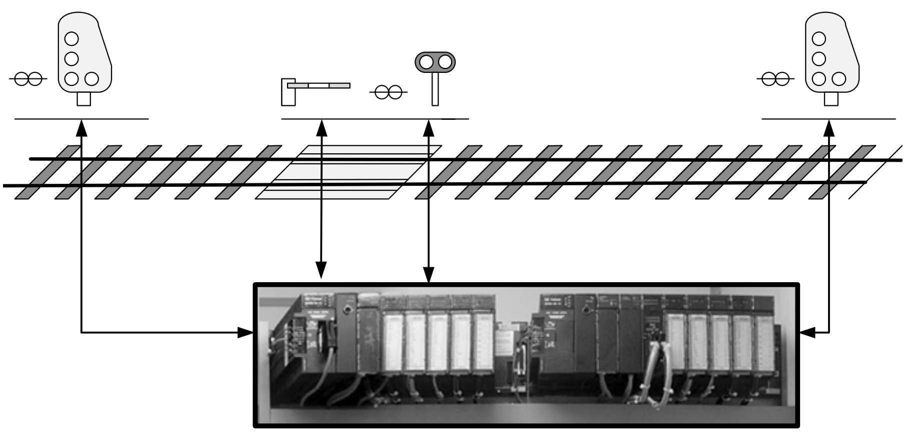


Fig. 1. Basic structure of the RASP-4 system [5]

The analysis of *RASP-4* system has been done in [2], [7]. The basic configuration of *RASP-4* systems includes: CPU, In/Out card and RSPAR14 card. For all of these components it was possible to estimate the failure rate λ:

- CPU - λ = 4.16E-05
- In card - λ = 1.21E-05
- Out card - λ = 9.48E-06
- RSPAR14 card - λ = 2.62E-05

The obtained value of failure rate λ for one channel equals:

$$\lambda_{kanał\backslash A} = \lambda_{kanał\ B} = 8{,}95 \cdot 10^{-5}\, h^{-1} \tag{4}$$

The global value of failure rate for the whole system equals:

$$\lambda_{RASP-4} = 5{,}96 \cdot 10^{-5}\, h^{-1} \tag{5}$$

2.2 New Solution of Cross Level Protection System

In connection with implementation of new technology for railway management and control systems, the wireless transmission can be used. The existing EU standards 50129.1, 50129.2 allow using of the public wireless standards to railway control and management, where necessary requirements of safety should be satisfied. Thanks to special cryptology method it is possible to secure wireless transmission. The simplified scheme of such a new system is presented in Fig. 2.

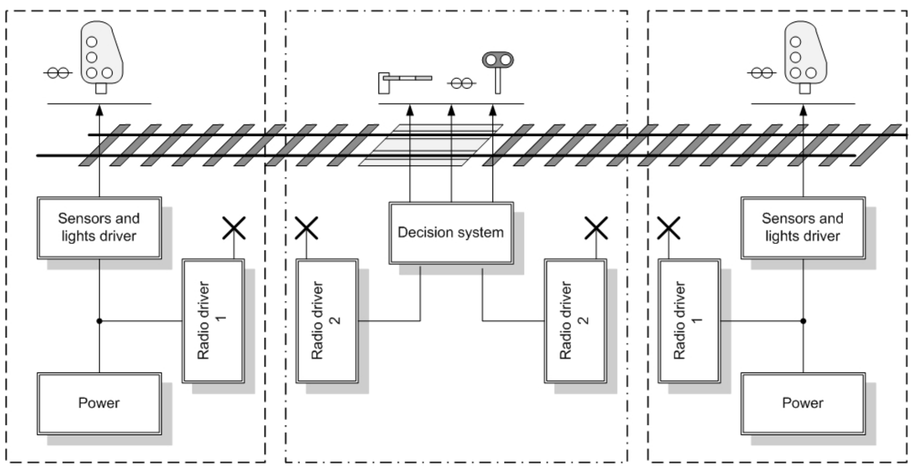


Fig. 2. Simplified scheme of the system

In the case in Fig. 2, there are not wire connections between some elements. Such project is realized by Kombud S.A. from Radom. There are redundant components in the system, which is indispensable for safety. The inside communication is realized by SPI (*Serial Peripheral Interface*) and Ethernet. Basing on Fig. 2, redundant structure is presented in Fig. 3.

The authors have done the basic analysis of safety of such a system (failure rate λ and *THR* rate) on the basis of technical documentation. The most important and necessary were the calculations of failure rate of electronic elements. For all electronic elements the G_F (*Ground, Fixed*) rate was assumed. The result of forecasting estimation is presented in Table 1.

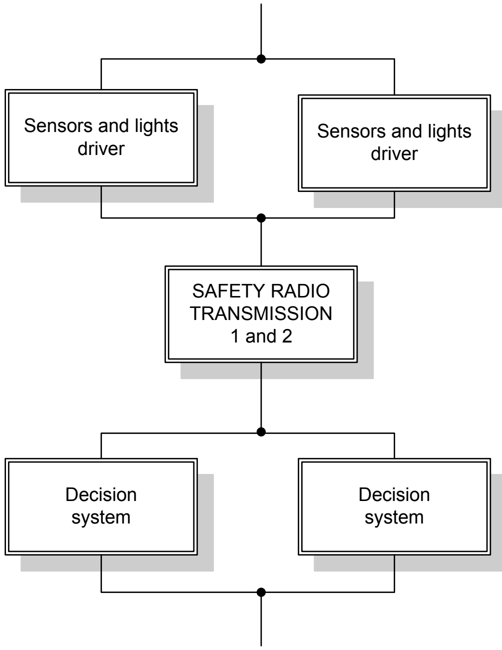


Fig. 3. Redundant structure

Table 1. The result of forecasting estimation

No.	Element	Failure rate value
1	Sensors and lights driver	8,8E-05
2	Radio driver 1	4,09558E-05
3	Radio driver 2	1,50242E-05
4	Decision system	2,54E-04

3 Verification

It was possible to assess the safety parameters on basis of exploitation result [5]. The quantity of failure in time (one range = 4500h) for six RASP-4 systems is presented in Fig. 4. For all systems in 27 000h there were only sixteen failures.

The estimated global value of failure rate λ equals:

$$\lambda = 7{,}40741 \cdot 10^{-5}\ \mathrm{h}^{-1} \tag{6}$$

In order to calculate the *THR* rate, the exploitation results were used, (6). Using (1) and assuming:

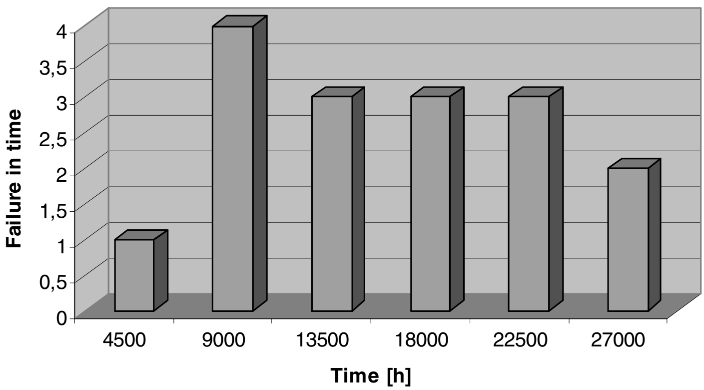


Fig. 4. The quantity of failures in time

- T = 500ms
- NT_{IN} = 1s
- NT_{OUT} = 1s

and because of time $NT_{IN} = NT_{OUT}$:

$$t_{d\,A} = t_{d\,B} = \frac{500\boldsymbol{ms}}{2} + 1\boldsymbol{s} = 1{,}25\boldsymbol{s} \tag{7}$$

Estimated $THR_{expl.}$ rate equals:

$$THR_{RASP-4(\exp l.)} = \frac{0{,}00007407}{\frac{1}{\left(\frac{1{,}25}{3600}\right)}} \cdot \frac{0{,}00007407}{\frac{1}{\left(\frac{1{,}25}{3600}\right)}} \cdot \left(\frac{1}{\left(\frac{1{,}25}{3600}\right)} + \frac{1}{\left(\frac{1{,}25}{3600}\right)} \right) = 3{,}8 \cdot 10^{-12} \tag{8}$$

Basis on forecasting estimation and assuming time t_d like in (7), *THR* rate equals:

$$THR_{RASP-4} = 5{,}56 \cdot 10^{-12} \tag{9}$$

For redundant components of new system, assuming time t_d like in (7), calculated values of *THR* equals:

$$THR_{New\_system(sensors)} = 5{,}37 \cdot 10^{-12} \tag{10}$$

$$THR_{New\_system(\text{decision system})} = 4{,}48 \cdot 10^{-11} \tag{11}$$

The radio transmission cards create autonomic safety transmission system and they are not included to THR estimation.

4 Conclusions

All results presented in the paper confirm high level of safety of the systems. Both *RASP-4* system and redundant components of new cross level protection system meet requirements concerned *THR* rate. For computer systems belonging to the highest level of safety, *THR* amounts: $10^{-9} \leq THR < 10^{-8}$. With reference to the *RASP-4* system, the verification of forecasting estimation was possible. In this case the results of forecasting estimation and exploitation results are similar. For new cross level protection system, on the actual stage of research, the forecasting estimation is rather approximated. The authors assumed the worst value of π rates for majority of electronic elements. After final acceptation, the electronic elements can be calculated on basis of producer's data. New cross level protection system is still developing and the verification of assumed safety parameters will be possible in the future. Presented analysis shows that for dissipated systems where some sub-systems are connected via safety transmission system the reliability of such systems may be better because their complexity is lower and in consequence the required values of *THR* are easier to satisfy.

References

1. Dąbrowa–Bajon, M.: The modelling in control railway. The railways problems, vol. (95). WKiŁ, Warszawa (1982) (in Polish)
2. Lewiński, A., Perzyński, T.: The verification of parameters of railway control systems on the basis of cross level protection system. In: LogiTrans. Conference, Szczyrk 2009 (in Polish)
3. Lewiński, A., Red.: The programmable railway control systems study. The scientist and research work IAiTT Technical University of Radom, 1999 (in Polish)
4. Lewiński, A.: The safety computer systems problems in railway transport application. Monograph Series, vol. 49. Technical University of Radom Publishing, Radom (2001) (in Polish)
5. KOMBUD S.A. Radom – documantation of cross level protection systems
6. Military Hand Book, Reliability Prediction of Electronic Equipment, USA Department of Defense (1991)
7. Perzyński, T.: The problems of safety of computer nets applied in the railway control. PhD Thesis – Technical University of Radom, Faculty of Electric Engineering and Transport, Radom (2009) (in Polish)
8. Polish Standard PN-EN 50129:2003

Belief Structures in Position Fixing

Wlodzimierz Filipowicz

Gdynia Maritime University, Faculty of Navigation
81-225 Gdynia, Poland
wlofil@am.gdynia.pl

Abstract. Mathematical Theory of Evidence extended for fuzzy environment proved to be universal platform for wide variety of new solutions wherever knowledge, uncertainty and imprecision are to be considered. Navigation is a discipline where all the mentioned factors are present. The paper has primary objective of introducing methods and mechanisms of the theory into the computation scheme while position fixing. The way of upgrading belief structure based on navigational aids indications is presented. Methods of dealing with pseudo belief structures are discussed.

Keywords: Mathematical Theory of Evidence, fuzzy reasoning, position fixing.

1 Introduction

In his previous papers, the author presented the concept of implementation of the Mathematical Theory of Evidence (MTE) in navigation. The theory proved to be flexible enough to be engaged in position fixing. Contrary to the traditional approach, it enables embracing knowledge and uncertainty into calculation scheme. Knowledge regarding position fixing includes: characteristics of random distributions of measuring errors, ambiguity and imprecision in obtained parameters. Uncertainty can be expressed by subjectively evaluated masses of confidence attributed to each of observations.

The theory of evidence deals with encoded facts and weights assigned to each of them. Weights can be assigned to any subset of the universe without assigning mass to particular elements of this subset. Consequently, the theory is able to cope with incomplete evidence and ignorance. Evidence can be delivered from different sources. It can be extracted from indications obtained with navigational aids. The theory provides scheme of combination of available evidence. Association operation is a kind of plausible reasoning. It enables obtaining numerical degrees of support to propositions relative to evidence at hand. In this respect informative context of initial sets of data is increased.

MTE operates on: belief and plausibility measures, encoded evidence, masses of evidence and belief structures. Structures used in navigation consist of fuzzy location vectors, which are specific events, and masses of evidence assigned to each of them. Structures are to be upgraded with reference to each indicated position, measured distance and/or bearing. Way of defining the structures is presented in the paper. Remarks on their normalization are also included.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 434–446, 2010.

2 Position Fixing in Terrestrial Navigation

Position fixing in terrestrial navigation usually engages distances and/or bearings. Distances circles intersect at several points in the vicinity of the ship position. Assuming measured distances as random variables, the true position of the ship is somewhere inside intersection area. It is navigator knowledge and experience that allow estimating the fixed position. The more accurate the measured distances the less is the area the better is estimation of the true position. Obviously, an experienced navigator is able to assess the acceptable dimensions of such area. Intersection region, greater than an average, causes rejection of the fix. Most common approach to analytical way of position fixing exploits the least square adjustment method.

Traditional way of position fixing engages the obtained indications. Their normal distribution is widely assumed and exploited in analytical method. Subjectively evaluated masses of credibility attributed to each of the measurements can be included in least square adjustment approach.

The proposed solution, based on MTE extended to fuzzy environment is more flexible and enables considering various random characteristic of the measured values, kind of distribution is important and can affect final solution. Empirical and theoretical distribution can be taken into account. Imprecision in accuracy estimation[1] can be considered. The approach delivers evaluation of selected position. Plausibility, belief (for appropriate definitions see [1] or [2]) and inconsistency values enable assessment of quality of the final solution.

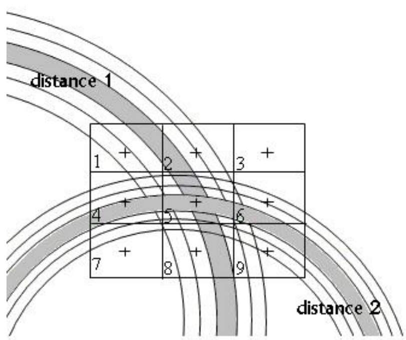


Fig. 1. Intersection area of two imprecise distances and search space grid with numbered cells. Centres of cells determine search space points.

The example of intersection area of two distances with search space grid is shown in Fig. 1. Very much like in traditional way in order to make a fix using MTE one has to explore intersection area of a distances and/or bearings. The suggested approach enables to include knowledge into calculations. It is assumed that the measured distance is a random value governed by normal distribution. Measurements characteristics are presented in Table 1. Subjective assessment of each measurement is also included.

[1] In navigation handbooks, for example in [6], one can read that mean error of measuring distance with radar is within ±1% to ±1.5%.

Table 1. Facts and knowledge regarding taken distances

considered factors	values
distance d_1	80 cables
standard deviation σ_{d1}	1.0 cable
standard deviation interval $[\sigma^-_{d1}, \sigma^+_{d1}]$	[0.8, 1.2] cables
degree of confidence assigned to the first measurement	0.6
distance d_2	60 cables
standard deviation σ_{d2}	0.75 cables
standard deviation interval $[\sigma^-_{d2}, \sigma^+_{d2}]$	[0.6, 0.9] cables
degree of confidence assigned to the second measurement	0.8

Grid cells centers, treated as search space points are to be located with respect to each measured values. Beside each of the obtained distances the ranges are established. The probability of the true distance or bearing within each range is assumed to be known. Assuming normal distribution of measured distances and/or bearing ranges may be selected based on standard deviation (σ_d) available for each navigational aid. It is known that starting from measured value within intervals equal to standard deviation probability of embracing a true distance can be calculated. Therefore the deviation is a primary factor in establishing ranges beside each indicated value – see Fig. 2. As it is seen from the figure three, ranges with imprecise limits at each side of the measured value were established. Ranges are named with *a, b, c* at right side and *a', b', c'* at left side of the indicated distance (see also stripes in Fig. 1). Heights of rectangles used for marking selected ranges reflect decreasing probabilities of containing the true distance. Probability that the true distance is behind $d \pm 3\sigma_d$ is very close to zero.

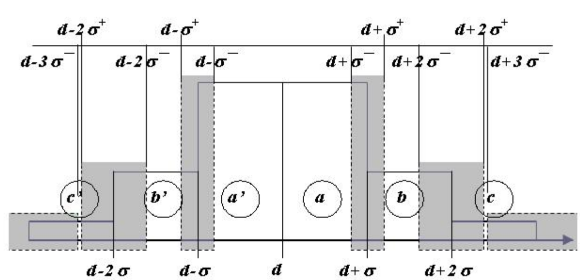


Fig. 2. Measured distance *d*, limits of selected ranges $d\text{-}2\sigma$, $d\text{-}\sigma$, $d+\sigma$, $d+2\sigma$. Limits of transition zones between pairs of ranges: *c' & b'* $\rightarrow d\text{-}2\sigma^+$, $d\text{-}2\sigma^-$; *b' & a'* $\rightarrow d\text{-}\sigma^+$, $d\text{-}\sigma^-$; *a & b* $\rightarrow d+\sigma^-$, $d\text{-}\sigma^+$; *b & c* $\rightarrow d+2\sigma^-$, $d\text{-}2\sigma^+$.

Unfortunately, the standard deviations cannot be treated as precise crisp values. Alternatively, they are imprecise interval valued $[\sigma^-_d, \sigma^+_d]$. In [6] one can read that mean error for distance taken with variable marker varies from ±1% to ±1.5%, and for bearings is between ±1° and ±2.5° provided medium class modern radar is used. Therefore interval valued transition zones between proposed ranges are to be introduced. For zone between range *a* and *b* is an interval $[d+\sigma^-_d, d+\sigma^+_d] = [d+0.01d, d+0.015d]$.

Respective interval valued gap for bearings would be: $[\alpha+\sigma^{-}_{\alpha}, \alpha+\sigma^{+}_{\alpha}] = [\alpha+1, \alpha+2.5]$. It should be noted that widths of transition zones increase while moving away from the measured value. The zones can overlap in case of broad initial range.

Probabilities of the true distance or bearing within the selected ranges are respectively equal to: 0.02, 0.14, 0.34, 0.34, 0.14 and 0.02. It should be noted that distinguishing adjacent ranges *a* and *a'* is not justified from the approach principles point of view. Thus both ranges will be considered jointly with the true value probability of 0.68.

Search space points are to be located with reference to measured distances. It is necessary to calculate their belonging to selected ranges. In case of any point situated inside transition zone its membership is split between adjacent ranges or even a few ranges in case of overlapping zones. In order to calculate membership within each of ranges appropriate functions have been suggested by the author in his previous paper [4]. Example membership function for range *b* is presented in Fig. 3.

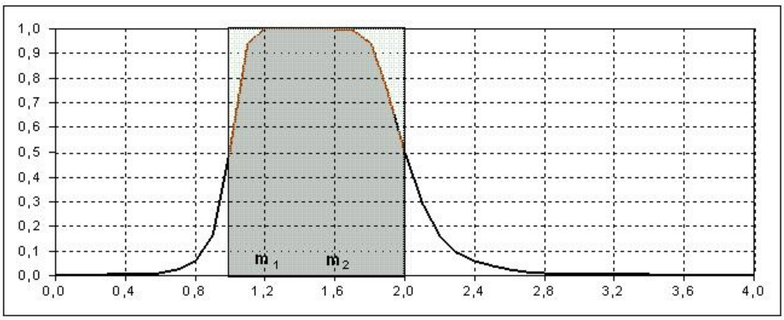


Fig. 3. Membership function for range *b* (standard deviation σ=1), m_1 and m_2 are limit abscissas for which function is equal to one. At borders of the range function is equal to 0.5. For given set of points {0.4, 0.8, 1.5, 2.4} their respective grades of membership within range *b* are {0, 0.06, 1, 0.06}.

Fig. 3 presents membership function or possibility distribution [7] expressing fuzzy belonging to a range of [1, 2]. Using relative scale, or assuming standard deviation equal to one, it is valid for fuzzy proposition "located within range *b*". The range is limited by $[\sigma, 2\sigma]$.

$$\mu_A(x;m_1,m_2,\alpha,\beta)=\begin{cases} L\left(\dfrac{m_1-x}{\alpha}\right) & if\ x\le m_1 \\ 1 & if\ m_1<x<m_2 \\ R\left(\dfrac{x-m_2}{\beta}\right) & if\ x\ge m_2 \end{cases} \tag{1}$$

$$L(x)=P(x)=\frac{1}{1+|x|^4} \tag{2}$$

Membership function presented in Fig. 3 was obtained using formula 1 ($\alpha = 0.4$, $\beta = 0.8$) with final engaging expression 2 to get a fuzzy figure of *L-P* kind [8]. Constants m_i represent limits of x-coordinates for which membership is equal to one. Factor α determines shape of the left slope and β of the right one of the membership function. The smaller are these factors the steeper are borders of the function.

Table 2 presents fuzzy locations for the scheme shown in Fig. 1. Each of the centers of the grid cells are located with reference to observed distances. Location factors were obtained using membership functions of a type presented in Fig. 3.

Table 2. Search space points locations with respect to the two distances

point	reference to distance 1		reference to distance 2	
	location	m. degree	location	m. degree
1	*c'*	1	*outside*	-
2	*b* & *c*	0.8 & 0.2	*outside*	-
3	*outside*	-	*outside*	0
4	*outside*	-	*a* & *b*	0.8 & 0.2
5	*a*	1	*a*	1
6	*outside*	-	*b* & *c*	0.3 & 0.7
7	*outside*	-	*outside*	-
8	*b'* & *c'*	0.6 & 0.4	*outside*	-
9	*outside*	-	partly in *c'*	0.7

m. degree – stands for membership degree within given range(s)
b & *c* – means that particular point belongs to two ranges *b* and *c*
0.8 & 0.2 – means that the point memberships within left and right areas (in this example *b* and *c*) are: 0.8 and 0.2 respectively
outside – stands for locations outside range *c* or *c'*

From the second row and the third column in Table 2 one can read that degree of location of point number 2 inside range *b* is equal to 0.8 and its presence within range *c* is estimated as 0.2. It explains the term "fuzzy location", which means that points belong, to some extent, to each of the ranges. In case of point located beyond ranges *c'* or *c* term *outside* is used. Since probability that the true distance or bearing is so far away from the measured one is very close to zero these ranges are neglected during calculations.

Table 3 shows belief structures related to position fixing with two distances. Facts gathered in Table 2 enable to define major parts of these structures. Encoded locations are called location vectors; their elements are the degrees of search space points belonging to the selected ranges with respect to each of the measured values. Since *a* range is treated jointly, five location vectors are within single belief structure, each related to particular range. Location vectors are supplemented with all one set, which expresses uncertainty. Mass attributed to this vector shows lack of confidence to the particular measurement. Thanks to this value all observations can be differentiated, usually in subjective manner.

All location vectors have assigned mass of confidence. Masses values are calculated based on probability attributed to particular range and on complement of uncertainty assigned to particular observation. Initial probabilities are multiplied by factors 1-$m(\mu_{in})$. It should be noted that the sum of all masses within a single belief structure is to be equal to one. Masses of evidence are called exact beliefs contrary to total beliefs that are values attributed to given set of propositions [1]. In considered case total belief determines credibility of representation a fixed position by each of search space points. Exact beliefs in location within given ranges contribute to the total belief in considering particular point as a fixed position. Apart from belief function plausibility one is even more important while making a fix [3], [4].

Table 3. Two belief structures enabling position fixing with two imprecise distances

	location vectors									*m(..)*
μ_{1a}	{0	0	0	0	1	0	0	0	0}	0.41
$\mu_{1b'}$	{0	0	0	0	0	0	0	0.6	0}	0.085
μ_{1b}	{0	0.8	0	0	0	0	0	0	0}	0.085
$\mu_{1c'}$	{1	0	0	0	0	0	0	0.4	0}	0.01
μ_{1c}	{0	0.2	0	0	0	0	0	0	0}	0.01
μ_{1n}	{1	1	1	1	1	1	1	1	1}	0.40
μ_{2a}	{0	0	0	0.8	1	0	0	0	0}	0.55
$\mu_{2b'}$	{0	0	0	0	0	0	0	0	0}	0.11
μ_{2b}	{0	0	0	0.2	0	0.3	0	0	0}	0.11
$\mu_{2c'}$	{0	0	0	0	0	0	0	0	0.7}	0.015
μ_{2c}	{0	0	0	0	0	0.7	0	0	0}	0.015
μ_{2n}	{1	1	1	1	1	1	1	1	1}	0.20

μ_{ix} - stands for a fuzzy set related to *i-th* observation and *x* range
μ_{in} - stands for all one set expressing uncertainty related to *i-th* observation
m(..) - stands for masses of evidence assigned to location vectors

3 Integration of Various Navigational Systems

While at the open sea, a few indications delivered by various navigational aids are often available. Usually, the indicated points are spread over a certain area. An indication is a random value that is situated in the vicinity of the true position of the ship. Indications should remain mutually independent [7]. Functional principles of the systems involved are to be different. As an example we can consider GPS and LORAN which exploit various media.

Example of three indications and search space grid is shown in Fig. 4. The area of a fixed position can be easily determined once random distributions and independence of all available points are valid. The area is a rectangle that covers extreme values of coordinates.

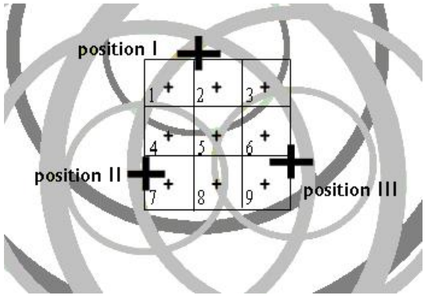


Fig. 4. Positions indicated by three navigational systems

Navigator's task is to make a fixed position of the ship based on indicated points, using his skill and experience. Like it is done for distances and/or bearings there are facts and knowledge one is to consider once fix is made. Table 4 shows facts and knowledge regarding obtained indications, which are assumed as random points normally distributed around the true position.

Table 4. Facts and knowledge regarding obtained indications

considered items	example values
position indicated by system 1	(x_1, y_1)
standard deviation of indications σ_1	1.0 cable
standard deviation interval $[\sigma^-_1, \sigma^+_1]$	[0.6, 1.4] cables
degree of confidence assigned to the first system	0.6
position indicated by system 2	(x_2, y_2)
standard deviation σ_2	0.75 cables
standard deviation interval $[\sigma^-_2, \sigma^+_2]$	[0.6, 0.9] cables
degree of confidence assigned to the second system	0.7
position indicated by system 3	(x_3, y_3)
standard deviation σ_3	0.5 cables
standard deviation interval $[\sigma^-_3, \sigma^+_3]$	[0.2, 0.8] cables
degree of confidence assigned to the third system	0.7

Assuming normal distribution of indicated positions, areas around the obtained point may be selected basing on standard deviations (σ_i) obtained for each navigational aid. Starting from indicated point areas with different probability of the true position can be established. As for the measured distances or bearings, standard deviation is a primary factor in establishing areas around indicated position. In this case three circular areas are distinguished. Areas are marked with *a, b, c*. Probability that the fix is behind a radius of $3\sigma_d$ is very close to zero.

Unfortunately, standard deviations cannot be treated as precise crisp values. Alternatively, they are imprecise interval valued $[\sigma^-_i, \sigma^+_i]$. Interval valued transition regions between proposed areas are to be introduced. For regions between consecutive pairs of areas they are rings limited by two radiuses $[\sigma^-_i, \sigma^+_d]$, $[2\sigma^-_i, 2\sigma^+_d]$ and $[3\sigma^-_i,$

$3\sigma^{+}_{d}$]. Interval values multiplied by an increasing integer figures result in widening of the ranges. Fig. 5 presents selected areas and intermediate ranges, membership functions are also diagrammed. It is shown that widths of the transition regions increase while moving away from the indicated point.

Fig. 4 presents positions indicated by three navigational aids. In order to infer on the true position of the ship one has to establish areas with known probability regarding fixed position. Probabilities of the true position locations within each of the areas related to an indicated point are: 0.68, 0.27 and 0.04. The figures are obtained assuming that distribution of positions is governed by Gauss function and the same functions can be applied to both coordinates.

Statement of facts related to determined search space and given constellation of indicated points is included in Table 5. The table contains nine rows, each devoted to a single search space point. Points are numbered as shown in Fig. 4. Every point is situated reference to the obtained indications. Locations are identified by area names and degrees of memberships also called as membership grades. Complete set of data regarding particular indication is embraced in two adjacent columns.

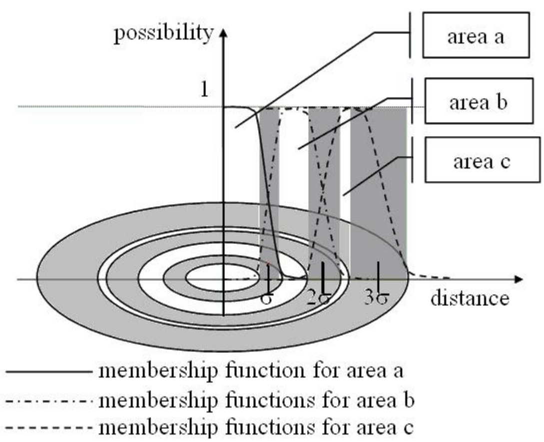


Fig. 5. Indicated position placed in the center of coordinate system with circular areas around. Areas are limited by radiuses: σ, 2σ and 3σ, their borders are treated as imprecise interval values.

Facts and knowledge regarding each of the indicated positions enable to create believe structures. Table 6 shows three belief structures enabling position fixing based on three systems indications. Encoded facts gathered in Table 5 were grouped by selected areas with respect to each indication. Each group creates fuzzy location vectors. Belief structure also contains masses of credibility assigned to location sets. In order to calculate masses the initial probability attributed to the particular range were multiplied by complements of the uncertainty assigned to each indication. Doubtfulness is expressed subjectively and is supposed to be fuzzy valued. Herein crisp values are used.

Table 5. Search space points locations with respect to the indicated positions

	reference to pos. I		reference to pos. II		reference to pos. III	
	location	m. degree	location	m. degree	location	m. degree
1	*a*	1	*b*	1	*b* & *c*	0.2 & 0.8
2	*a*	1	*b*	1	*c*	1
3	*a* & *b*	0.7 & 0.3	*b* &*c*	0.8 & 0.2	*b* & *c*	0.9 & 0.1
4	*a* & *b*	0.1 & 0.9	*c*	1	*b*	1
5	*a* & *b*	0.6 & 0.4	*a* &*b*	0.5 & 0.5	*a* & *b*	0.5 & 0.5
6	*b*	1	*b*	1	*a*	1
7	*b*	1	*a*	1	*b*	1
8	*b*	1	*b*	1	*a* & *b*	0.5 & 0.5
9	*b*	1	*b*	1	*a*	1

location – means position within selected area(s)
m. degree – stands for membership degree within given area(s)
a & *b* – means that particular point belongs to two areas *a* and *b*
0.7 & 0.3 – means that the point memberships within left and right areas (in this example *a* and *b*) are: 0.7 and 0.3

Table 6. Three belief structures enabling position fixing with imprecise indications

	location vectors									*m(..)*
μ_{1a}	{1	1	0.7	0.1	0.6	0	0	0	0}	0.408
μ_{1b}	{0	0	0.3	0.9	0.4	1	1	1	1}	0.168
μ_{1c}	{0	0	0	0	0	0	0	0	0}	0.024
μ_{1n}	{1	1	1	1	1	1	1	1	1}	0.400
μ_{2a}	{0	0	0	1	0.5	0	1	1	0}	0.476
μ_{2b}	{1	1	0.8	0	0.5	1	0	0	1}	0.196
μ_{2c}	{0	0	0.2	0	0	0	0	0	0}	0.028
μ_{2n}	{1	1	1	1	1	1	1	1	1}	0.300
μ_{3a}	{0	0	0	0	0.5	1	0	0.5	1}	0.476
μ_{3b}	{0.2	1	0.9	1	0.5	0	1	0.5	0}	0.196
μ_{3c}	{0.8	0	0.1	0	0	0	0	0	0}	0.028
μ_{3n}	{1	1	1	1	1	1	1	1	1}	0.300

μ_{ix} - name of a fuzzy set related to *i-th* position and *x* area
μ_{in} - name of all one set expressing uncertainty
m(..) - stands for masses of evidence assigned to location vectors

4 Combination and Normalization of Belief Structures

The Theory of Evidence requires that mass of evidence assigned to null set is to be zero. Violation of this requirement mean that some mass is assigned to null vector what indicates inconsistency occurrence. Sum of all masses assigned to sets is to be equal to one. Another requirement regarding structures with fuzzy sets stipulates that the sets are to be normal. The set is normal when its highest grade is equal to one. Assignment for which above requirements are not observed is pseudo belief structure and is to be normalized.

Tables 3&6 contain belief structures related to observations indicated by various navigational aids. Each structure consists of location vectors supplemented with all its uncertainty set. First structure contains assignment to μ_{1c} that is null vector. Apart from this row, subnormal fuzzy location vectors can be found. Thus initial sets are pseudo belief structures that are to be converted into their normal states. Two normalization procedures are widely used one was proposed by Dempster, another one by Yager. At first, both of them considered crisp events. Further extensions for fuzzy environment were suggested by Yager [9]. Although it is quite often that many authors refer to them using original methods inventor names.

Normalization procedures are quite different in two aspects. Masses of inconsistency in Dempster approach increase weights attributed to not null sets. In Yager proposal the masses increase uncertainty. In case of subnormal sets Dempster suggested division by highest grade, Yager proposed adding complement of the largest grade to all elements of the set. Results of subnormal fuzzy set conversion to its normal state using Yager and Dempster methods are different. Note that fuzzy set is a location vector containing fuzzy memberships of a search space points within selected ranges or areas. Thus Dempster transformation causes that points with not null locations increase their memberships. In Yager normalization all considered points gain some degrees of membership. Unfortunately it may adversely affect selected position.

Table 7. Two fuzzy sets, their combination and normalization

	location vectors									*m(..)*
μ_1	{0	0.8	0	0	0	0	0	0.6	0}	0.408
μ^Y_1	{0.2	1	0.2	0.2	0.2	0.2	0.2	0.8	0.2}	0.168
μ^D_1	{0	1	0	0	0	0	0	0.75	0}	0.476
μ_2	{0	0	0	0.67	0	1	0	0	0}	0.196
$\mu^Y_{\mu1\wedge\mu2}$	{0	0	0	0.2	0	0.2	0	0	0}	0.196
$\mu^D_{\mu1\wedge\mu2}$	{0	0	0	0	0	0	0	0	0}	0.300

μ^Y_1 - result of normalization of fuzzy set μ_1 using Yager method
μ^D_1 - result of normalization of fuzzy set μ_1 using Dempster method
$\mu^Y_{\mu1\wedge\mu2}$ - result of combination of fuzzy sets μ_1 and μ_2 normalized with Yager method
$\mu^D_{\mu1\wedge\mu2}$ - result of combination of fuzzy sets μ_1 and μ_2 normalized with Dampster method

Table 7 presents combinations of two fuzzy location vectors μ_1 and μ_2 representing membership of the search space points within certain areas. The two mentioned areas do not embrace the same points; in this meaning they are disjunctive and their intersection is empty. Therefore, the product of masses attributed to both combined vectors should be assigned to empty vector what means inconsistency occurrence. This case is missed as result of using Yager normalisation. Result of combination is not null subnormal vector hence it is transferred into its normal state but uncertainty is not modified. The last is done provided converted vector is a null set. Contrary to this approach, inconsistency can be easily detected while using Dempster method. In this case the result of combination of fuzzy belongings to two disjunctive areas is a null set and inconsistency occurs. In original Dempster attitude this occurrence should be forgotten after updating of all not null assignments. Any assignment to null sets is to be zeroed. In considered application inconsistencies are mainly related to constellation of given set of positions. The more they are dispersed the higher the inconsistency is and fix quality gets worse. Therefore, the sum of inconsistencies masses are to be calculated and analyzed, none of occurrence should be missed. Normalization procedure should not contribute to misidentification of inconsistencies. Thus, the extended Dempster normalization is to be proposed for nautical applications. The extension covers recording and analyzing of all inconsistency occurrences. It should be emphasized that inconsistency can occur for overlapping areas provided that there is no search space within the common region. In other words sizes of the whole grid as well as single cell can also affect inconsistency value.

Table 8. Results of reasoning on fixed position based on combination of belief structures from Table 6

	1	2	3	4	5	6	7	8	9
$pl^D(..)$	0.21	0.29	0.25	0.33	0.63	0.32	0.32	0.41	0.32
$bel^D(..)$	0.01	0.02	0.00	0.00	0.15	0.00	0.00	0.01	0.00
uncertainty = 0.18									
$pl^Y(..)$	0.39	0.46	0.44	0.49	0.73	0.51	0.48	0.60	0.51
$bel^Y(..)$	0.01	0.02	0.00	0.00	0.07	0.00	0.00	0.00	0.00
uncertainty + inconsistency = 0.16									

$pl^D(..)$ - plausibility values obtained for Dampster normalization
$bel^D(..)$ - belief values obtained for Dampster normalization
$pl^Y(..)$ - plausibility values obtained for Yager normalization
$bel^Y(..)$ - belief values obtained for Yager normalization

In practical applications, belief structures are subject of combination [10]. Those presented in Table 6 were combined with Yager as well as Dempster normalization applied to intermediate and final results. Then, in the normalized final results support for proposition of representing fixed position of the ship are sought regarding each search space points. Table 8 contains calculated plausibility (*pl*) and belief (*bel*) values for considered search space determined by centres of the grid cells. The highest plausibility and belief values are assigned to point number 5. Since plausibility is considered as primary factor in position fixing [5] then this cell is pointed as representing the fixed position. The cell and its neighbourhood can be further explored in order to achieve required accuracy [4]. It should be emphasised that uncertainty is equal to 0.18 for Dempster transformation. Inconsistency recorded during association plus uncertainty is 0.16 for Yager normalization.

Table 9 contains comparison of Dempster and Yager normalizations taking into account the practical aspects presented in the first column. If should be noted that position fixing engages fuzzy location vectors therefore Dempster normalization should be recommended. The approach also proved to dominate in inconsistency cases detection as well as in fact that final solution remained unaffected by combination condition. For more navigational usage of belief structures and their normalization see previous paper by the author [3].

Table 9. Dempster versus Yager approach

	Dempster normalization	Yager normalization
way of modification of masses assigned to not null sets	increased by a factor calculated using inconsistency values	remained unchanged
final uncertainty	depend on initial masses assigned to all one sets	uncertainty is increased by inconsistencies
modification of membership grades	general shape of location vectors is preserved, null grades remain unchanged	null grades of location vectors gain some membership
ability to detect all inconsistency cases	guaranteed	doubtful
recommended for	belief structures with fuzzy events	belief structures with binary events
not recommended for	belief structures with binary events and high inconsistency	belief structures with fuzzy events and high inconsistency
computational complexity	medium	low
high inconsistency and its effect on final solution	not observed	high inconsistency may adversely effect final solution

5 Conclusions

New method of position fixing in navigation is proposed. The method uses MTE concept, which enables reasoning on position fixing based on taken measurements or available indication. It was assumed that measured values or indicated points are random variables with known distribution. The true distance or bearing is somewhere in the vicinity of the obtained one. To select fixed location a few ranges were introduced in the vicinity of each observation. Probability levels assigned to each range are calculated based on features of given distribution. Standard deviation or another parameter of the distribution is assumed to be an imprecise data known within a certain range.

In order to indicate fixed position of the ship search points are to be defined. It was assumed that centers of grid cells determine the space. The grid is spanned over potential position area. Search space points located with membership functions within established ranges create fuzzy sets called location vectors. Fuzziness results from imprecise interval valued estimations of standard deviations.

Fuzzy location vectors supplemented with the one expressing uncertainty compose one part of belief structure. Another part embraces masses of initial believes assigned to location vectors and uncertainty. Complete belief structure is related to each of available measurements. Mass assigned to uncertainty expresses subjective assessment of measuring conditions. One has to take into account: radar echo signature, height of objects, visibility and so on to estimate measurement quality. Fuzzy values such as poor, medium or good can be used instead of crisp figures. Imprecise masses values engage different way of calculation.

Belief structures are combined. During association process search space points within common intersection region are selected. Result of association is to be explored for reasoning on the fix. All associated items are to be taken into account in order to select final solution.

The Theory of Evidence requires that mass of evidence assigned to null set is to be zero and fuzzy sets are to be normal. Assignment for which above requirements are not observed is pseudo belief structure and is to be normalized. Pseudo belief structures can occur at the structures preparation stage as well as during association process. Usually, null sets are results of combination of two ranges or areas without common search space points. The occurrences indicate abnormality in computation that might result from extraordinary erroneous measurements and/or wrongly adjusted search space. Therefore, all null assignment cases are to be recorded and analyzed. Two normalization procedures proposed by Dempster and Yager are widely used. Converting procedures are quite different in two aspects. Masses of inconsistency in Dempster approach increase weights attributed to not null sets. In Yager proposal the masses increase uncertainty. In case of subnormal sets Dempster suggested division by highest grade, Yager proposed adding complement of the largest grade to all elements of the set. The latter causes that Dempster approach should be perceived as superior in case of position fixing.

References

1. Barnett, A.J.: Computational Methods for a Mathematical Theory of Evidence. In: Proceeding of Seventh International Joint Conference on Artificial Intelligence Vancouver, pp. 868–875 (1981)
2. Denoeux, T.: Modelling Vague Beliefs using Fuzzy Valued Belief Structures. Fuzzy Sets and Systems 116, 167–199 (2000)
3. Filipowicz, W.: On Normalization of Belief Structures. In: Mikulski, J. (ed.) Advances in Transport Systems Telematics, pp. 73–81. WKL, Warsaw (2009)
4. Filipowicz, W.: Belief Structures and their Application in Navigation. Methods of Applied Informatics 3, 53–82 (2009)
5. Filipowicz, W.: Mathematical Theory of Evidence and its Application in Navigation. In: Knowledge Engineering and Expert Systems, pp. 599–614. EXIT, Warsaw (2009)
6. Jurdzinski, M.: Foundations of Marine Navigation. WAM, Gdynia (2008)
7. Lee, S.E.: Fuzzy and Evidence Reasoning. Physica-Verlag, Heidelberg (1995)
8. Rutkowski, L.: Methods and Techniques of the Artificial Intelligence. PWN, Warsaw (2009)
9. Yager, R.: On the Normalization of Fuzzy Belief Structures. International Journal of Approximate Reasoning 14, 127–153 (1996)
10. Yen, J.: Generalizing the Dempster-Shafer theory to fuzzy sets. IEEE Transactions on Systems, Man and Cybernetics 20(3), 559–570 (1990)

Application of TTCN-3 for Testing of Railway Interlocking Systems

Zbigniew Łukasik[1] and Waldemar Nowakowski[2]

[1] Technical University of Radom, Faculty of Transport,
Malczewskiego 29, 26-600 Radom, Poland
z.lukasik@pr.radom.pl
[2] Z.A. KOMBUD S.A., Wrocławska 7, 26-600 Radom, Poland
waldemar.nowakowski@kombud.com.pl

Abstract. The assurance of required safety integrity level in the field of railway signalling is essential. The most commonly used verification, validation and testing methods are not exhaustive and can miss errors. This paper describes an alternative verification technique based on the TTCN-3 (Testing and Test Control Notation). The TTCN-3 is an internationally standardized testing language for formal defining of test scenarios and their implementation.

Keywords: Railway interlocking system, verification, testing, TTCN-3.

1 Verification and Testing of Railway Interlocking Systems

Railway interlocking systems are safety-critical and must therefore be developed according to the highest safety integrity level (SIL-4). The concept of a safety integrity level has different meanings when applied to different entities. The system safety integrity level is a number that indicates the required degree of confidence that a system will meet its specified safety features, where safety is the freedom from unacceptable levels of risk. This means that a system safety integrity level that is suitable for its required reliability will be allocated to the given system. A software safety integrity level carries a slightly different meaning. In this context, it is the classification number that determines the techniques and measures that have to be applied in order to reduce residual software faults to a tolerable level. Since it is applicable to software, and since the software is a part of the system, it is derived from the system safety integrity level [3].

A safety design for the railway interlocking system begins at the specification phase [7]:

- System Safety Plan,
- Safety Concept,
- System Safety Requirements Specification,
- Software Quality Assurance Plan,
- Software Verification Plan,
- Software Validation Plan,
- Software/Hardware Integration Test Plan.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 447–454, 2010.

Verification of software is troublesome. Various techniques may be used, with varying difficulty and results. The CENELEC standard lists a number of criteria in different areas of software quality assurance.

- Static analysis (e.g. boundary value analysis, data flow analysis, error guessing, etc.),
- Dynamic analysis and testing (e.g. test case execution, equivalence classes and input partition testing, etc.),
- Performance testing (e.g. avalanche- and stress testing, response timing and memory constraints, performance requirements, etc.).

We have chosen the TTCN-3 to implement our solution for dynamic testing of railway interlocking systems. The research work relating to the subject conducted in the TT-Medal project (Tests & Testing Methodologies with Advanced Languages) are now on a preliminary level [1] and [2].

2 Testing and Test Control Notation Version 3

The TTCN-3 is a language with syntax and operational semantics standardized by ETSI (European Telecommunication Standards Institute). It was originally developed for real-time testing telecommunication systems. Typical application areas are the testing of protocols, services, modules, and application programming interface. The TTCN was first standardized in 1992. In the year 2000 ETSI published the version TTCN-3. Since then the TTCN-3 is maintained through a well-defined change request process handled by ETSI and it is currently available in version 4.1.1, published in 2009.

The TTCN-3 is standardized in a multi-part deliverable that comprises ten parts:

- (ES 201 873-1) Core Language (CL),
- (ES 201 873-2) Tabular Presentation Format (TFT),
- (ES 201 873-3) Graphical Presentation Format (GFT),
- (ES 201 873-4) Operational Semantics,
- (ES 201 873-5) Runtime Interface (TRI),
- (ES 201 873-6) Control Interface (TCI),
- (ES 201 873-7) The use of ASN 1,
- (ES 201 873-8) The IDL to TTCN-3 Mapping,
- (ES 201 873-9) Using XML schema with TTCN-3,
- (ES 201 873-10) Documentation Comment Specification.

The TTCN-3 test system introduces a conceptual collection of entities that manage tests execution and interpret or execute compiled TTCN-3 code [6]. Figure 1 gives an overview of TTCN-3.

The textual core notation, also called “core language”, is at the heart of TTCN-3. The core notation is similar to conventional programming languages but provides additional features dedicated to testing such as built-in data matching, test verdicts, timer handling, or the concurrent execution of test components. The core notation can be represented using other formats than the textual one (e.g. TFT, GFT). Furthermore, the core notation provides interfaces to description languages (e.g. ASN.1, XML).

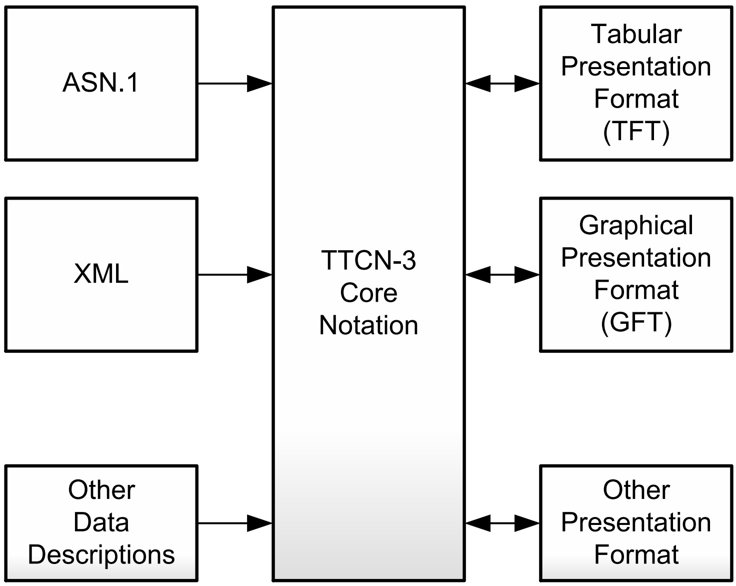


Fig. 1. Core language and various presentation formats

The authors of the article carried out research on description language ASN.1 (Abstract Syntax Notation One) in the field of railway signalling. The research has shown that the use of ASN.1 significantly simplifies designing of communication and, at the same time, enables interface standardization of computer railway control systems [4].

A TTCN-3 test system can be conceptually defined as a collection of interacting entities where each entity corresponds to a particular aspect of functionality in a test system implementation. The TTCN-3 test system architecture consists of the following components (Fig. 2):

- **TE** – TTCN-3 Test Executable entity is responsible for the realization of a test system (interpretation or execution),
- **TM** – Test Management entity is responsible for the overall management of a test system,
- **TL** – Test Logging entity is responsible for the logging of information about the execution of a test system,
- **CD** – Coding/Decoding entity is responsible for the encoding and decoding of TTCN-3 values into bit strings suitable to be sent to the SUT,
- **CH** – Component Handling entity is responsible for the implementation of distribution and communication between the distributed entities,
- **SA** – SUT Adaptor entity is responsible for the communication with the SUT,
- **PA** – Platform Adaptor entity is responsible for the realization of timers and their interaction with the TE,
- **SUT** – System Under Test.

The module control part is a set of test components, ports, specific interfaces and the SUT. Every test system has one Main Test Component (MTC) and optional Parallel Test Components (PTC). The MTC is created automatically at the beginning of a test case execution, whereas PTCs can be created, started and stopped dynamically (i.e. during test's execution). The test case execution shall end when the MTC terminates. The communication between test components (using ***connected***) and between the components and the test system interface (using ***mapped***) is achieved via communication ports.

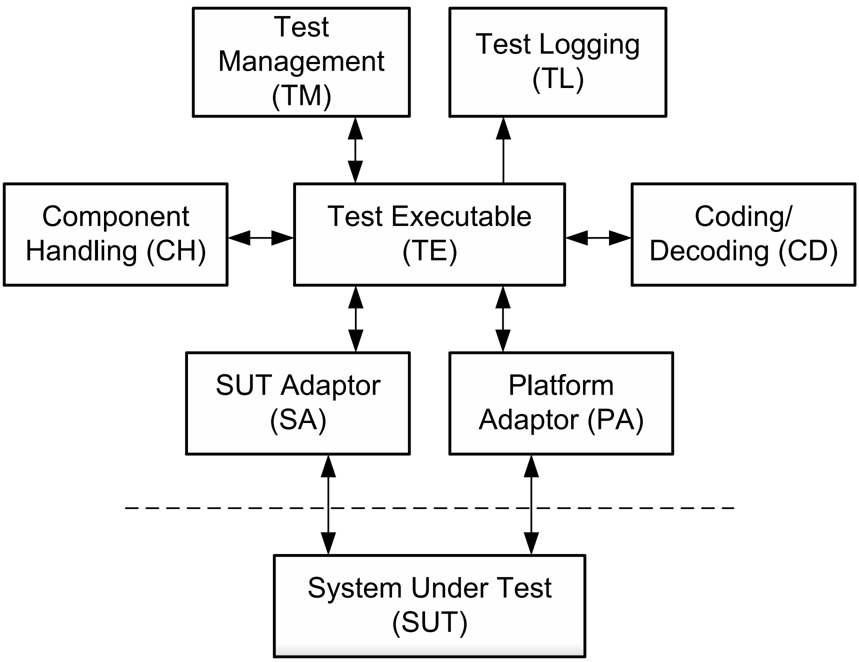


Fig. 2. Structure of a TTCN-3 test system

A TTCN-3 test system has two interfaces: the abstract and the real test system interface. The abstract test system interface is a collection of ports defining the abstract interface to the SUT. Port mapping between a test component and the abstract test system interface is a mere name translation determining the means of referring to communication streams. Components, ports and the abstract test system interface are application independent, whereas the real test system interface is application specific, implementing the real interface to the SUT [5].

3 Abstract Syntax Notation One

ASN.1 may be used with TTCN-3 modules as syntax of data types. The syntax of ASN.1 is very similar to the ones used in most of the programming languages. However it lacks keywords related to the execution flow, such as loops or conditional statements. The language contains a few built-in data types and a set of rules, which enable individual creation of new, more complex data types, as well as assignment of constant values to some elements contained in data types defined this way. Abstract types related to each other are grouped together in modules which are represented in the TTCN-3 via their names. Simplifying, a module may be introduced as the following structure [4]:

ModuleIdentifier **DEFINITIONS** ::=
BEGIN
    ModuleBody
END
where: **DEFINITIONS**, **BEGIN**, **END** are keywords for ASN.1,
ModuleIdentifier is the name that identifies the module,
ModuleBody –the body of the module.

Data types in ASN.1 standard may be divided into basic and constructed ones. Basic types can hold values from predefined set of values. The group of basic types includes: BOOLEAN, NULL, INTEGER, ENUMERATED, REAL, BIT STRING, OCTET STRING, OBJECT IDENTIFIER, CHARACTER, and STRING. Constructed types are formed by combining basic types into structures. The group of constructed types includes structural types: SEQUENCE, SEQUENCE OF, SET, SET OF and choice types: CHOICE, ANY. By means of types mentioned one may define different data structures.

The application in TTCN-3 data structures defined of ASN.1 provides a lot of advantages, such as:

- abstractness - the notation uses abstract syntax which enables description of data structures irrespective of their further representation,
- structurality - the notation enables decomposition of data structures into smaller units called modules. This enables preservation of greater clarity, which in turn facilitates verification and modification of data structures,
- formal definitions - the syntax and semantics of the notation have a complete and formal definition. The formal model allows using analytical theory in order to prove correctness.

4 TTCN-3 Railway Interlocking System Tester Example

In order to evaluate the use of TTCN-3 language as a tool for dynamic testing of railway interlocking systems, a simple experiment was conducted by means of “Loong Testing” developed by the University of Science and Technology of China [8] (Fig. 3).

The “Loong Testing” is an application that has a parser and a compiler of modules, created in the TTCN-3 language. It allows defining data structures in the TTCN-3 and in ASN.1 [4]. As a result of tests compilation there are modules in C++ language and adequate DLL libraries. It enables to verify immediately the defined test in the TTCN-3 and present the results in a textual format.

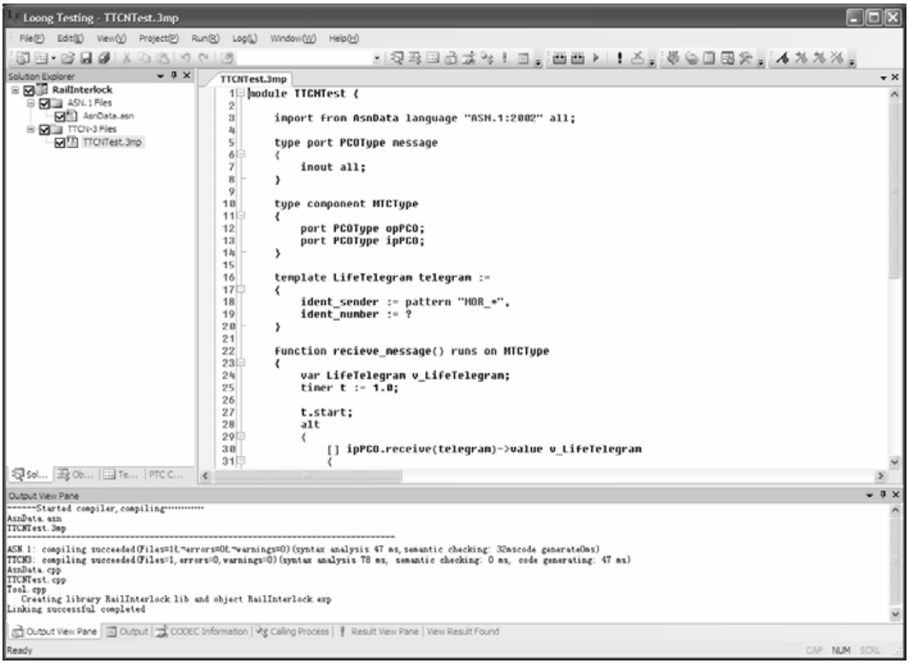


Fig. 3 “Loong Testing” tool.

A simple experiment, using the TTCN-3, was conducted. The object of research was a railway interlocking communication interface. The example of ASN.1 data types for “LifeMessage” is presented below:

```
AsnData DEFINITIONS ::=
BEGIN
  LifeMessage  ::= SEQUENCE  {
    -- Interlocking data
    ident-sender IA5String (SIZE(1..8)),
    ident-number INTEGER (1..65535)
  }
END
```

The messages can be sent from SUT to the test system which will verify them. The data exchange is conducted with the use of “opPCO” and “ipPCO” ports:

```
import from AsnData language "ASN.1:2002" all;
type port PCOType message {
    inout all;
}
type component MTCType {
    port PCOType opPCO;
    port PCOType ipPCO;
}
template LifeMessage message := {
```

```
        ident_sender := pattern "MOR_*",
        ident_number := ?
    }
    function recieve_message() runs on MTCType {
        var LifeMessage vLifeMessage;
        timer t := 1.0;
        t.start;
        alt
        {
              ipPCO.receive(message)->value vLifeMessage {
              log(vLifeMessage);
              setverdict(pass)
              }
              [] ipPCO.receive {
                          log("Failed message");
                          setverdict(fail)
              }
              [] t.timeout {
                    setverdict(fail)
              }
        }
    }
```

The list of alternatives “alt”, defined in the “receive_message” function of component “MTCType”, is used in this case. The logging information about the execution of a test system is saved in the TL entity (Fig. 2).

5 Conclusions

A railway interlocking is a part of a traffic control system that is responsible for the safety of the system. This paper describes a case study on testing of railway interlocking systems with the TTCN-3. A framework for testing of railway interlocking systems has been developed, starting with the mapping of railway safety requirements to TTCN-3 test cases and proceeding with the implementation of a railway TTCN-3 test system. The successful application of TTCN-3 for testing of railway interlocking systems is a step towards automation and standardization of testing processes in such systems. The use of those methods enables functional safety verification of railway interlocking systems.

References

1. Blom, S., Ioustinova, N., Van de Pol, J., Rennoch, A., Sidorova, N.: Simulated Time for Testing Railway Interlockings with TTCN-3. In: Grieskamp, W., Weise, C. (eds.) FATES 2005. LNCS, vol. 3997, pp. 1–15. Springer, Heidelberg (2006)
2. Calame, J.R., Goga, N., Ioustinova, N., Van de Pol, J.: TTCN-3 Testing of Hoorn-Kersenboogerd Railway Interlocking. In: Proceedings of the Canadian Conference on Electrical and Computer Engineering, CCECE 2006, Ottawa, Canada (2006)
3. Hedberg, K., Elestedt, F.: Safety-critical Communication Controllers for Railway Signalling in Public Networks. Chalmers University of Technology, Göteborg (2008)

4. Łukasik, Z., Nowakowski, W.: ASN.1 notation for exchange of data in computer-based railway control systems. Transport Problems, Silesian University of Technology, Gliwice 4(2) (2009)
5. Mäki-Asiala, P.: Reuse of TTCN-3 Code, vol. 557. Espoo VTT Publications (2005)
6. Willcock, C., Deiß, T., Tobies, S., Keil, S., Engler, F., Schulz, S.: An Introduction to TTCN-3. John Wiley & Sons Ltd., New York (2005)
7. EN 50129:2003- Railway applications - Communication, signalling and processing systems - Safety related electronic systems for signalling
8. University of Science and Technology of China, http://ttcn.ustc.edu.cn

Maritime Intelligent Transport Systems

Zbigniew Pietrzykowski

Maritime University of Szczecin, Institute of Marine Navigation,
Waly Chrobrego 1-2, 70500 Szczecin, Poland
z.pietrzykowski@am.szczecin.pl

Abstract. The concept of intelligent transport systems, developed since 1980s, includes all modes of transport. Relevant developments in road transport are the most advanced. Recently water transport, particularly maritime transport, has gained more attention in this respect, in connection with building and further developments of maritime intelligent transport systems. The present state and directions of development of telematic systems in maritime transport will be discussed, with a focus on marine navigation.

Keywords: telematics, intelligent transport systems, maritime transport, navigation.

1 Introduction

Advancements in information and communications technologies (ICT) create wider opportunities for improving performance of maritime and all other modes of transport. Such possibilities are found in construction and extending the capabilities of intelligent transport systems. Research on these systems started in Japan and the USA in the late 1980s, aimed at building national intelligent road transport systems.

Modern information technologies have also been present in maritime transport for some time now. Due to the specific and mostly international character of sea transport, the implemented solutions have to take account of the legal instruments in force, including relevant IMO resolutions. This means that new solutions require approval of each country concerned, consultations with neighbouring states as well as, in many cases, approval of international bodies. We can suppose that was one of the reasons for less interest shown by particular countries in the building of national maritime intelligent transport systems.

The existing solutions, comprising such systems as Automatic Identification System (AIS), Vessel Traffic Services (VTS), Vessel Traffic Management Systems (VTMS), Vessel Traffic Management and Information Systems (VTMIS), Long Range Identification and Tracking system (LRIT), SafeSeaNet, recently executed Maritime Navigation and Information Services (MarNIS) project, as well as the concepts of Motorways of the Sea (MOS), e-Navigation and e-Maritime show that there is a need for setting forth guidelines for and architecture of maritime intelligent transport systems.

The above mentioned solutions, projects and concepts as well as experience gained in creating intelligent transport systems in road transport may be useful in the maritime sector.

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 455–462, 2010.

2 Maritime Transport System

Maritime transport is one of the major links in the global transport chain. Like in other modes of transport, those who render transport services, apart from economic aspects, have to take into account the safety of people, means of transport (vessels), cargo and the environment. Maritime transport services are provided by the maritime transport system.

Transport system elements are objects taking part in the process of moving cargoes and people and objects related with the movement process. These include: seaports, sea-going ships as cargo carrying vehicles, waterways, equipment and traffic arrangements including regulations for traffic safety and control. People participating in the provision of transport services are major elements of the system.

The transport system approached dynamically corresponds to the dynamic transport process. This process can be divided into sub-processes taking place within the system and those taking place outside and forcing action in the transport system. The sub-processes inside the system, in turn, can be divided into decision and technological sub-processes. These processes are determined by technical, economic and organizational constraints as well as those imposed by the environment.

The system tasks result from the major function of transport services: carriage of people and/or cargo. These tasks include organization, control and supervision of the transport process, which in turn, consists of such operations as loading / embarkation, carriage by ship, unloading / disembarkation. From another perspective these operations are services offered by the system and rendered by its sub-systems. These include, but are not limited to:

- freight and fleet management,
- vessel traffic management,
- safety management and damage control management,
- management of information for shipowners, marine agents, vessel commanders, travellers, administrative bodies and others.

The major participants of the transport process are:

- shipowners (to optimize transport service operations of their fleets),
- companies responsible for the organization and execution of un/loading processes,
- operators responsible for the organization and supervision of vessel traffic within port approaches and basins (to efficiently handle ingoing and outgoing traffic: assist in collision avoidance, control movements of ships, minimize their waiting times before arrival and departure),
- safety and security services (safety and damage control management – organization and command of rescue operations),
- navigators on watch (to complete a sea voyage according to shipowner's instructions and, at the same time, observe the rules of safe navigation, regulations in force and instructions of operators responsible for the organization and supervision of vessel traffic).

Solutions implemented at present and actions undertaken, such as projects to be realized and concepts under development, are aimed at consistent enhancement of maritime transport quality and efficiency, increasing its attractiveness and competitiveness, and the improvement of the transport process link in the global chain. Advancement of information exchange processes and better co-ordination of parties involved in the transport process is of key importance. Such opportunities are possible owing to dynamically progressing ICT developments.

3 Telematic Systems in Maritime Transport – The Present State

Systems using IT and telecommunications solutions have found applications in maritime transport for many years, both on ships and in land-based centres.

Today numerous electronic navigational and communications technologies and services are available or at the stage of development. These comprise Global Navigation Satellite Systems (GNSS), Automatic Identification System (AIS), Electronic Chart Display and Information Systems (ECDIS), Integrated Bridge Systems/ Integrated Navigation Systems (IBS/INS), Automatic Radar Plotting Aids (ARPA), radio navigation, Long Range Identification and Tracking (LRIT) systems, Vessel Traffic Services (VTS) and the modified generation of Global Maritime Distress and Safety System (GMDSS) - which can provide automatically the navigators on watch and those ashore with the necessary information they require.

These systems are mainly used on ships (ARPA, GNSS, AIS, ECDIS, GMDSS, IBS/ INS) and by land-based centres responsible for the vessel traffic safety, supervision and management, including state bodies: maritime offices and harbour master offices subject to them, Port State Control, border and customs services, crisis management centres (apart from the above mentioned, LRIT, SaveSeaNet and other specialized systems, such as VTS) and port authorities. Some of them, AIS in particular, are used by shipowners, port agents, forwarders, cargo consignees and passengers.

Satellite navigational systems (e.g. GPS, Glonass, Compass, Galileo) provide autonomous geo-spatial positioning with global coverage. GNSS allows to determine object position (longitude, latitude, and altitude) with high accuracy, using signals from satellites.

ARPA is a computer-aided radar data processing system, which presents a current situation of a ship. Its basic functions are automatic or manual acquisition of targets, determination of own ship's movement parameters and those of acquired targets (digital read-out providing course, speed, range, bearing, closest point of approach), signalling dangerous situations, the ability to perform trial maneuvers, including changes of the course and/or speed.

The AIS system performs automatic data exchange between ships (also used as an additional source of information for collision avoidance by ships) and ship identification by e.g. shore-based vessel traffic services. Among other functions, AIS automatically and continuously (with appropriately frequent data transmissions) sends ship's static and dynamic data, such as voyage data (draft, dangerous goods carried, port of destination, estimated time of arrival) as well as short safety messages, updated as need arises.

ECDIS is a navigational information system, capable of presenting graphically selected information from built-in electronic navigational chart (ENC) with overlaid information from other navigational devices and systems (log, radar, ARPA, echo sounder, GNSS, AIS). Its major functions are as follows: presentation of all information contained in an ENC, necessary for safe and efficient ship navigation, position, own ship movement and acquired targets approach parameters, generating alarms and warnings concerning the presented data, routing, track control and voyage data recording.

Today these shipboard systems are in most cases combined and incorporated into an integrated navigational bridge (IBS).

The LRIT is a reporting system for ships which are required by IMO regulations to report to their flag administration. LRIT information is sent to competent maritime administrations and search and rescue services entitled to receive the information, upon request, through a system of National, Regional, Co-operative and International LRIT Data Centres.

SafeSeaNet is a platform of data exchange between EU member state administrations, enabling tracking ships underway, managed by the European Maritime Safety Agency (EMSA).

GMDSS, Global Maritime Distress and Safety System, encompasses a set of safety procedures, radio and satellite communication systems aimed at ensuring the safety of navigation and prompt and effective alerting to accidents at sea, transmitting information critical for vessels' safety and providing communication during search and rescue operations.

VTS systems are information systems of vessel traffic centres supervising and controlling vessel movements. Initially VTS facilities featured radars and ARPA devices installed on ships. Today they are dedicated systems implementing, among others, functions of ARPA, ECDIS and AIS.

To enhance their effectiveness, VTS systems are expanded with management modules, supporting the operator in making decisions (Vessel Traffic Management Systems). VTMS combines three basic VTS services: information, navigational assistance and traffic organization, into one management system. The extension of system functions mainly applies to in-port traffic management.

Vessel Traffic Management and Information Systems (VTMIS) are systems created by merging single VTS and VTMS systems to make them co-operate and thus supervise larger navigational areas. One characteristic feature of these systems is their use of specialized navigational equipment and systems which support the processes of information acquisition, processing and presentation.

In 2002 work started on creating a European Vessel Traffic Monitoring and Information System (VTMIS) [1]. This system includes all sea areas of EU members and consists of their national VTMIS. VTMIS is composed of:

- Ship Reporting Systems (SRS),
- Maritime Assistance Services (MAS) with places of refuge,
- Long-range Identification and Tracking System (LRIT)
- national centres of SafeSeaNet – a European Platform for Maritime Data Exchange.

It is assumed that "The European Vessel Traffic Monitoring and Information System should help to prevent accidents and pollution at sea and to minimise their impact on the marine and coastal environment, the economy and the health of local communities. The efficiency of maritime traffic, and in particular of the management of ships' calls into ports, also depends on ships giving sufficient advance notice of their arrival " [1].

The operation of the above systems is based on various communication systems:

- means of terrestrial communication
- means of sea-based communication via digital VHF, AIS, WiMax or UMTS,
- satellite communication.

One may expect that the ship-to-shore and intership communication will be developing towards a wider use of satellite communications and extended coverage of terrestrial communication systems.

4 Concepts of Telematic Systems in Maritime Transport

The forums of the European Parliament and IMO propose and develop concepts aimed at better utilization of ICT in maritime transport. These include Motorways of the Sea, e-Navigation, e-Maritime, Maritime ITS. Complementary to these are concepts of e-cargo, e-custom and others, concerning all modes of transport.

The idea of Motorways of the Sea emerged in 2001 [2]. Implemented within the EU, it is intended to improve transport processes in short-sea shipping and intermodal transport. A Motorway of the sea comprises the infrastructure and organization in two EU ports. Operations along a sea motorway should be based on electronic systems of logistics management of passenger and cargo traffic, safety systems and simplified administrative and customs procedures.

Four sea transport corridors are regarded as essential for the EU: 1) Motorway of the Baltic Sea; 2) Motorway of the Sea of western Europe; 3) Motorway of the Sea of south-east Europe; Motorway of the Sea of south-west Europe. The basis for developing the concept of motorways of the sea for the EU were projects implemented under the EU Interreg III B program.

The idea of e-Navigation is a concept of wider use of ICT in maritime navigation, developed in the IMO forum. To this end a Correspondence Group on e-Navigation was established within the IMO in 2006. The definition in [3] reads that "e-Navigation is the harmonized collection, integration, exchange, presentation and analysis of maritime information onboard and ashore by electronic means to enhance berth-to-berth navigation and related services, for safety and security at sea and protection of the marine environment". The aim is to develop an overarching accurate, safe, secure and cost-effective system with the potential to provide global coverage for vessels of all sizes. e-Navigation will consist of [4]:

- electronic navigational charts,
- positioning systems – use of satellite and terrestrial radionavigation system services (back-up systems),

- shipboard identification and information exchange systems capable of transmitting data on a vessel, its route, manoeuvring parameters, status, etc.,
- systems of ship-to-ship and ship-to-shore radio-communication,
- systems integrating shipboard and shore centre displays, enabling selection of information (priorities) and alerting in critical situations.

Before the concept becomes a real system, the methodology of creating such systems has to be followed: developing a conceptual model, specifying requirements and the construction of a logical model. These make up a basis for designing the system and its subsequent implementation. The mentioned Correspondence Group on e-Navigation coordinates work in this respect.

The concept of e-Maritime has been evolving since 2001 within EU countries [2]. The objective of e-Maritime initiative is to promote "coherent, transparent, efficient and simplified solutions in support of cooperation, interoperability and consistency between member States, sectors, business and systems involved in the European Transport System" [5]. The idea of e-Maritime encompasses policies, strategies, capabilities, including services and systems, within the maritime sector, supporting an efficient and sustainable waterborne transport system throughout Europe, fully integrated within the transport logistic chains.

The e-Maritime concept was developed within the framework of the EC co-funded 6th Framework project MarNIS (Maritime Navigation and Information Services) [6], which aimed at the improved exchange of information and provision of services and the required infrastructure to meet the requirements placed on both the authority and business level. The stakeholders include ships, shipowners, operators and agents, and on the other hand shore-based entities, including maritime authorities (e.g. Search and Rescue (SAR), coastal and port), related authorities (e.g. customs and immigration) and commercial parties within the port sector. MarNIS is based on the SafeSeaNet system and concepts of Maritime Information Management and Maritime Operational Services. The project proposes to put an interface for information exchange in the system of the so called single window, formed by management information systems: National Single Window (state administrations) and Port Single Window (port authority). The "single window" concept permits to access information with a single entry point, which simplifies processes of data collection, integration, verification and presentation. Other components of the system are, *inter alia*, Port Commercial Community Systems, Vessel Traffic Management in ports, VTS and Port Operations and Approach Decision Support System.

The e-Maritime concept will be developed by the intended construction of an integrated EU system providing services by electronic means at various levels of the transport chain, laid down by the EU in [7]. It is envisaged that this system will integrate e-Maritime with e-Freight, e-Customs and intelligent transport systems (ITS).

5 Maritime ITS

The aim of creating intelligent transport systems is to improve the efficiency of transport operations by, among other things, implementation of information technologies. Work on the development of ITS started in Japan and the US, where they concerned road transport, with a focus on increasing the safety of road traffic. For instance, an official document of the Japanese government [8] of 1999 defines nine areas constituting major directions of development and implementation of ITS, such as: Advances in navigation systems, Electronic toll collection systems, Assistance for safe driving, Optimization of traffic management, Increasing efficiency in road management, Support for public transport, Increasing efficiency in commercial vehicle operations, Support for pedestrians, Support for emergency vehicle operations.

ITS systems, built in individual countries, are based on the ITS architecture, a set of general principles, taking account of the specifics of a given country and its transport system, determining a framework for designing these systems. The architecture also defines the system functions and corresponding physical subsystems and their interfaces, communication requirements for the provision of efficient exchange of information, and, finally, it identifies standards for interoperability within a country and region. For instance, the main components of ITS architecture in Japan and the USA [9] (road transport) are: 1) units providing user services (User services); 2) logical architecture; 3) physical architecture. The logical architecture presents a functional view of the ITS user services. The physical architecture partitions the functions defined by the logical architecture into classes, and at a lower level, subsystems. Additionally, the Japanese ITS comprises standardization area (standardization candidate areas), while the ITS USA includes a set of tools and technologies enabling ITS users to achieve desired objectives (market packages) and the architecture documentation (Navigating the Architecture Documentation). An analysis of ITS architectures in other countries shows that they reflect a similar philosophy and differences are slight. The architecture concerns road transport, therefore it cannot be directly adapted to the maritime sector. It can, however, provide directions while defining a Maritime ITS architecture.

Actions undertaken at present, taking into account the integration of all modes of transport – including maritime transport – aim at:

- rationalization of transport system/s operation,
- counteracting accidents,
- counteracting traffic congestion,
- counteracting environment pollution.

As far as the maritime sector is concerned, ICT implementation follows many paths. EU’s concept of E-Maritime is one idea that may provide a basis for developing an open Maritime ITS architecture. E-Maritime fulfills the above mentioned aims. It defines users, system boundaries and its components: e-Maritime Reference Applications for administration and business domains, e-Maritime Support Platform and e-Maritime Strategic Framework. Conceptual and logical models elaborated under the MarNIS project make up a basis for developing implementation models. E-maritime concept, developed for the EU, may correspond to a national ITS for the maritime sector.

Discussion on Maritime ITS was undertaken at the 17th ITS World Congress in Stockholm, Sweden 2009, and will be continued at the ITS World Congress in South Korea in 2010. Designers of ITS architecture

6 Conclusions

The existing maritime telematic systems herein discussed and related concepts and projects in the stage of development show that information and communications technologies find increasingly wider applications in maritime transport. These concepts, particularly e-Maritime developed by the European Union, provide grounds for defining a general architecture of Maritime ITS, and its implementation in other regions of the world.

References

1. Directive 2002/59/EC of the European Parliament and of the Council of 27 June 2002 establishing a Community Vessel Traffic Monitoring and Information System, http://europa.eu/legislation_summaries/transport/waterborne_transport/l24243_en.htm
2. European Commission (EC) White Paper - European transport policy for 2010, COM (2001) 370, http://europa.eu/documentation/official-docs/white-papers
3. IMO, NAV 53/13. Development of an E-Navigation Strategy. Report of the Correspondence Group on enavigation, submitted by the United Kingdom. Sub-Committee on Safety of Navigation. In: International Maritime Organization, London, April 20 (2007)
4. Weintrit, A., Wawruch, R., Specht, C., Gucma, L., Pietrzykowski, Z.: Polish Approach to e-Navigation Concept. Coordinates III(6), 15–22 (2007)
5. European Commission (EC) Green Paper -Towards a future Maritime Policy for the Union, COM, 275 (2006), http://europa.eu/documentation/official-docs/green-papers
6. Maritime Navigation and Information Services MarNIS, Final Report (2009), http://www.marnis.org
7. European Commission (EC) Communication:Strategic goals and recommendations for the EU's maritime transport policy until 2018, COM, 8 (2009) , http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0400:FIN:en:PDF
8. System architecture for ITS in Japan (1999), http://www.its-jp.org
9. Executive Summary. Architecture Development Team. Federal Highway Administration, US Department of Transportation (2002), http://www.its.dot.gov

Analysis of Road Traffic Accident Rate in the Slovak Republic and Possibilities of Its Reduction through Telematic Applications

Alica Kalašová and Zuzana Krchová

University in Žilina, Faculty of Operation and Economics of Transport and Communications, Department of Road and Urban Transport, Univerzitná 1, 01026 Žilina, Slovakia
{Alica.Kalasova,Zuzana.Krchova}@fpedas.uniza.sk

Abstract. There is a worldwide trend of increasing number of vehicles being newly introduced on an annual basis. On the road, there are great numbers of drivers with low discipline, aggressive driving habits and non-compliance with basic principles of responsible traffic behavior. As a consequence, the number of traffic accidents is rising, and so do their repercussions.

There are more than 350 people, who annually die as victims of transport accidents in Slovakia. In addition, there is even larger number of injured which is alarming. Financial and other costs of transport accidents are estimated at the level of several billions of Euros annually.

In our article, we would like to outline an analysis of the transport accidents in Slovakia as well as we would want to show telematic applications, which lead to decreasing the number of transport accidents.

Keywords: Intelligent transport system, accidents, safety.

1 Introduction

Efficient transport services are key from the perspective of European industry competitiveness. Transport contributes greatly to its growth, but also creates negative externalities, which are, for instance, estimated at 1.1% of the European GDP [9]. Overload of the roads constantly increases and deprives the GDP of more than 1%. In the field of safety on the roads, Europe in the White Paper on the transport policy has specified itself an ambitious aim of decreasing the death of people by 50% till 2010 as compared to 2001. In the meantime, the safety has increased significantly, but the situation is still not satisfactory.

The Slovak Republic as a full member of the European Union respects the recommendations of the European Commission in the sphere of road safety and is trying to accomplish them.

The development of transport in all its sectors is linked with the integration into the advanced countries and with developing of the society. Year by year the number of vehicles on the roads as well as of new drivers is increasing and a lot of negative effects are associated with this situation. The number of road accidents and their consequences increase as a result of missing conditions for the implementation of transport education, aimed at improving low discipline, aggressive driving, violation of

J. Mikulski (Ed.): TST 2010, CCIS 104, pp. 463–468, 2010.

fundamental duties and low legal awareness of drivers and other participants of the road traffic. [1], [4] and [7].

The traffic accident rate as a serious social problem requires a comprehensive and effective solution that shows features of a coordinated and aimed procedure by all stakeholders and institutions with a broad public support [5].

2 Analysis of Traffic Accident Rate in the SR

The year 2009 could have been assessed from the point of the consequences of road traffic accidents as one of the best since 1966. In 2009 we recorded 25,975 accidents. Overview of the last 10 years is shown in Table 1. [5] and [6].

Table 1. Development of Basic Indicators for the Last 10 Years

Year	The number of accidents	The number of fatal accidents	The number of seriously injured	The number of slightly injured	The damage in EUR
1999	55.683	647	2 684	8 782	97 226 981
2000	50.932	628	2 204	7 890	86 781 331
2001	57.258	614	2 367	8 472	105 099 184
2002	57.060	610	2 213	8 050	108 261 793
2003	60.304	645	2 163	9 158	126 574 670
2004	61.233	603	2 157	9 033	133 463 146
2005	59.991	560	1 974	8 516	132 683 275
2006	62.040	579	2 032	8 660	140 407 019
2007	61.071	627	2 036	9 274	149 828 170
2008	59.008	558	1 806	9 234	149 594 690
2009	25 975	347	1 400	7 127	101 824 120

3 Characteristics of ITS

Intelligent Transport Systems (ITS) bring information and communication technology into the transport infrastructure and vehicles to improve safety and to reduce overload of the roads, shorten transportation time and to decrease the fuel consumption. With the use of telematic applications we are increasing the safety of the road transport and the comfort of traffic with the use of new transport information and communication technologies.

New technologies for safety can be divided into:

- intelligent active safety systems;
- intelligent passive safety systems;
- next generation advanced systems for assistance to drivers,
- cooperating vehicle-vehicle and vehicle-infrastructure systems.

Technological applications can also be seen as the most important groups of instruments, which can contribute to greater road safety through [8]:

- more intelligent vehicles and safer, more performing mobility,
- more intelligent infrastructure;
- information and communication technologies.

3.1 The Future in the Telematic Applications as Support for Increased Safety

As it was mentioned above, the European Community decided to request member states to implement such measures that the number of traffic fatal accidents could be reduced by 50% till 2010, as compared to the state in 2001. Looking ahead, the expected reduction of the number of fatal accidents should continue until 2020 with the aim to reduce their number by about 75% of the state in 2001. We have to remind that according to current statistics 93% of traffic accidents are caused by a human fault. Based on the current trend it can be observed that without radical measures will not be possible to meet those commitments.

Assistance systems are the main challenge. They are based on the communication (data exchange) not only among vehicles themselves but also vehicles and infrastructure. These so called Intelligent Assistance Systems promise great benefits in the sphere of efficiency of transport systems and road safety. These benefits include mainly an increase in the capacity of the road network, a reduction of congestion and pollution, a shorter and more predictable time of driving, improvement in traffic safety for all participants of the road traffic, lower operational costs for vehicles, better organization and management of road networks.

Cooperative systems for the information transmission in real-time utilise the communication between vehicles (Vehicle-to-vehicle, V2V) and between a vehicle and infrastructure (Vehicle-to-Infrastructure, V2I). They hold the promise of major improvements in the efficiency of the transport system, improve safety for all road users and increase the convenience that the mobility provides. The work on cooperative systems started in Europe in the fifth and sixth Framework Programme. Consortium Car2Car has been set up in connection with the industry [9], which promotes a common progress of industry. The key prerequisites and the main objective of the Commission include the development of harmonized and interoperable system architecture, a common communications architecture that can meet the needs of public and private sector, as well as the availability of suitable frequency.

To this end, as a part of the specific action COMeSafety [9], which was funded by the 6RP, a Working Group on Communications Architecture was established. This working group coordinates and consolidates the work of European and national projects as well as other key initiatives, and represents an interface for standardization in ISO and other standardization authorities.

3.2 Tasks and Services of the V2V2I Communication System (Vehicle-to-Vehicle-to-Infrastructure)

It is clear that a reliable communication system which enables secure and reliable communication between a vehicle and the units of transport infrastructure, as well as between vehicles, could significantly contribute to increase the safety of traffic operations, the use of transport infrastructure and to increase the user comfort in the entire transport system. The scheme of V2V2I system is shown in Fig. 1, which presents individual types of roads V2V: Vehicle - vehicle (vehicle to vehicle), V2I: vehicle - infrastructure (vehicle to infrastructure) and DRC2I: traffic operate center - infrastructure.

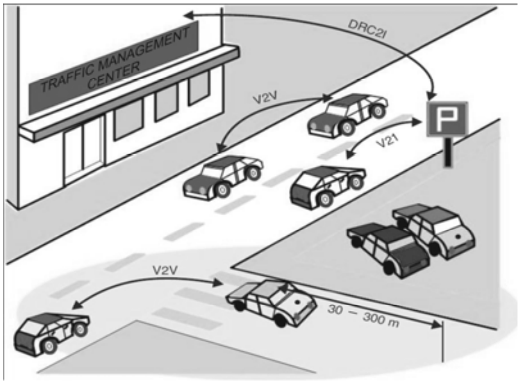


Fig. 1. V2V2I Communication System [3]

The communication system enables the provision of services, which directly relate to the road safety, as well as services, that are related to a better awareness of drivers, by selecting different fees, and to the ability to connect to public communication networks and so on.

Nowadays, we meet with the classification of services that are based on various criteria. It is suitable to divide services based on using a communication system V2V2I to the following three groups [2] and [3]:

1. Services directly related to safety such as: e.g. warning about collisions, about crossroads, about works on the road, about the presence of black ice and fog, about vehicle preference, etc.
2. Services related to the optimal use of transport infrastructure such as: better updating of navigation systems, data for the optimal speed in the case of a green wave, dynamic routing, guidance to the free parking areas, etc.
3. Services related to the other applications such as remote vehicle diagnostics and updating the software of on-board computers, making the payments (road fees, parking fees, payments for fuel, food) and so on.

4 Slovakia and ITS

The Slovak Republic faces a lot of transport problems, which result not only from the uncompleted transport infrastructure, but which concern many areas such as for example the transport safety, impact of transport on the environment or quality of service, which were not sufficiently resolved in the past. In 2005, the Government of the Slovak republic accepted the Transport Policy of the Slovak Republic till the year 2015 by means of Resolution No. 445, which defines a global and several specific objectives that include concrete measures in the transport in Slovakia to ensure a sustainable development of mobility, to secure long-term continually increasing movement needs of the society in the required time and quality and with parallel reduction of negative effects on the environment. Draft action plan is prepared in several areas, which are aimed at:

- optimization of the use of the latest road data and data access,
- the synergy of combined applications and services in the area of commercial transport,
- data security and protection of personal and commercial data,
- strengthening the influence of public authorities in the area of ITS,
- a framework for integration and coordination of programs,
- a framework for the promotion of diversity and acceptance of ITS.

The support program of the development of intelligent transport systems – the National System of Traffic Information represents a comprehensive solution of intelligent transport systems based on information and communication systems and technologies in the road transport in Slovakia. It is oriented to the use of a unified system environment for the collection, processing, sharing, distribution and use of transport information in concrete information, operate and telematic applications. In January 2009 a resolution of the Government was adopted on the project of ITS development support - establishment of the National System of Transport Information (NSDI) by the end of 2013.

Basic system requirements:

- to minimize the creation of congestion,
- to increase the efficiency of traffic, that is expressed by time savings,
- to increase mobility and quality of transport services,
- to create space for an efficient multimodal transport,
- to make available real-time information for passengers, carriers and users of communications,
- to improve the productivity of commercial activities of transport processes in the society,
- to reduce the energy consumption,
- to increase the quality of the environment.

5 Conclusion

Recently, the growth of road transport is a phenomenon associated with the development, which manifests in a significant growth of negative impacts of transport on the environment, growth in the congestion in conurbations and growth in traffic accidents, similar both in developed countries, and in conditions of the Slovak Republic.

The Forthcoming Action Plan of ITS is aimed at:

- decreasing the transport congestion by 25% and increasing the quality of travel,
- increasing the transport safety by 25% and thereby contributing to the overall European goal to reduce the number of people deaths by 50%,
- reducing CO_2 emissions by 10 %, mainly in urban areas.

The Acknowledgments

This work was supported by the project: Centre of excellence for systems and services of intelligent transport, ITMS 26220120028 University of Žilina, Žilina, Slovak republic

Project Part-Financed by the European Union
European Regional Development Fund

Grant VEGA 1/0398/08 Negative externality reduction in road transportation.

References

1. Konečný, V., Poliak, M.: Trh hromadnej osobnej dopravy a jej financovanie - 1. vyd. vedeckej monografie - Žilina : Žilinská univerzita (2009); 176 s., AH 12,00, VH 12,47 - ISBN 978-80-8070-999-0
2. Meitzner, R.: Kick of Meeting SeVeCom. Lausanne, Swiss (2006)
3. Miček, J: Komunikačný systém pre zvýšenie bezpečnosti a efektívnosti, Automatizace, ročník 51 číslo (Apríl 4 2008)
4. Ondruš, J.: Methodology of Acceptance Feasibility Survey of Urban Road Pricing. In: Transcom 2009, University of Žilina (2009), ISBN 978 -80-8070-692-0
5. Prezídium policajného zboru: Štatistický prehľad nehôd v cestnej premávke, SR 2008, Bratislava (Apríl 2009)
6. Sedláčková, A. Novák, A.: Situácia na trhu leteckej dopravy v Európe včera a dnes /. In: Ekonomicko-manažérske spektrum. Vedecký časopis Fakulty prevádzky a ekonomiky dopravy a spojov Žilinskej univerzity v Žiline. - ISSN 1337-0839
7. Špalek, J. ,Kállay, F., Janota, A., Koncepcia eSAFETY v inteligentných dopravných systémoch, In: Horizonty dopravy, 2/2003, ISSN 1210-0978, str. 7-9
8. Car 2 car, Communication Consorcium, `http://www.car.-to-car.org`
9. European Commision, `http://www.ec.europa.eu`

Author Index