Part 2

TRAIL 5th Annual Congress 1999

Five Years "Crossroads of Theory and Practice"

TRAIL Anniversary Congress

December 1st, 1999
Kurhaus Hotel, The Hague/Scheveningen
The Netherlands
Proceedings

Five Years "Crossroads of Theory and Practice"
TRAIL Anniversary Congress

5th TRAIL Annual Congress 1999

Part 2

Proceedings

5th TRAIL Annual Congress 1999

Part 2
Proceedings

3rd TRAIL Annual Conference 1993

Paul 2
Five Years "Crossroads of Theory and Practice"
TRAIL Anniversary Congress

Proceedings

5th TRAIL Annual Congress 1999

Part 2
Editor : Piet H.L. Bovy

The Netherlands TRAIL Research School
Delft University of Technology /
Erasmus University Rotterdam/
University of Groningen

Published and distributed by:
Delft University Press
P.O. Box 98
2600 MG Delft
Telephone: +31 15 278325
Telefax: +31 15 2781661
E-mail: DUP@DUP.TUDelft.nl

TRAIL Conference Proceedings Series N°. P99/5

The Netherlands TRAIL Research School 1999

ISBN 90-5584-034-3

© Copyright by The Netherlands TRAIL Research School. No part of this book may be reproduced in any form by print, photoprint, microfilm or any other means without written permission of the publisher: The Netherlands TRAIL Research School.

TRAIL is accredited by the Royal Netherlands Academy of Arts and Sciences and is a co-operative venture in which seven faculties of the Delft University of Technology participate: Civil Engineering and Geosciences (CEG), Design, Engineering and Production (DEP), Architecture (ARC), Information Technology & Systems (ITS), Aerospace Engineering (AES), Technology, Policy & Management (TPM), Applied Sciences and OTB Research Institute for Housing, Urban and Mobility studies; four faculties of the Erasmus University of Rotterdam: Economic Sciences (FEB), Business Administration (FBA), Social Sciences (FSS) and Law (FL), and one faculty of the University of Groningen: Behavioural and Social Sciences (BSS)
Contents

Part 2

Dynamic Traffic Management

The Free Speed Distribution of Drivers: Estimation Approaches
H. Botma

Modelling and Estimation of Dynamic Origin-destination Matrices
Ch.D.R. Lindveld

Architecture of an Agent Based Urban Intersection Control System
D.A. Roozemond

Traffic Control and Route Choice: Occurrence of Instabilities
H. Taale
H.J. van Zuylen

Vehicle Control Technology Impacts

Drivers' Stated Preferences for a Driver State Monitoring System
V.A.W.J. Marchau
E.J.E. Molin, S. van Hoytema, D. de Waard

Planning & Maintenance of Infrastructure

A Method to Reduce Excessive Noise Nuisance
A.A.J. Nederveen

Introducing Life Cycle Cost Management in the Railways: Opportunities and Drawbacks
A. Zoeteman

Integrated Land-use & Transport Modelling
The Structure of the Land-use Model
B. Zondag

Transport Policy Analysis

Policy Making in the Netherlands: The Limits of an Open Approach
M. Esselbrugge

The Role of Transport Costs in Goods Transport
H.A.C. Runhaar
B. Kuipers
Locked-in Logistics and Schiphol Airport
A First Step towards a Theoretical Framework
P.M.J. Warffemius

Seamless Multimodal Mobility

Modelling Pedestrians in Transfer Stations
W. Daamen

On the Optimisation of Buffer Times in Transport Networks
R.-J. van Egmond

Accessibility Service Levels as an Instrument to Improve the Accessibility of Special Activity Areas
A. van der Elst

Improving Punctuality and Transfer Reliability by Railway Timetable Optimization
R.M.P. Goverde

Travel Choice Behavior in Public Transport Networks
A Fuzzy Genetic Approach
S. Hoogendoorn-Lanser
S.P. Hoogendoorn

Influence of Delays on Departure Times in Air Traffic
G. Soto Y Koelemeijer
H. van der Steen

Designing a Multimodal Coding Scheme for PITA
R.J. van Vark

Appendix : The Netherlands TRAIL Research School
## Contents

### Part 1

**A Quasi-Variational Inequality Approach to Multi-User Class Dynamic Traffic Assignment**
M.C.J. Bliemer

**Modelling and Evaluation of Urban Freight Distribution Systems**
J.H.K. Boerkamps
A.J. van Binsbergen

**Traffic Scheduling: System Optimum versus User Equilibrium Models**
K. Koolstra

**Institutions and Change in Seaports**
P.W. de Langen

**Routing Vehicles On-line with Forecasted Release Times to Approach Off-line Performance**
J.R. van der Meer

**On-line versus Off-line Vehicle Control for Internal Transport Systems**
J.R. van der Meer

**Larger Stop Spacing in Urban Transit Networks: Why?**
R. van Nes

**An Optimization Approach to Railway Timetabling**
L.W.P. Peeters

**Design Approach for Terminals in Automated Underground Transportation Systems. The OLS case.**
B.A. Pielage

**Demand-responsive Personal Transport: The Maastricht case 'Vervoer op Maat'**
R.M. Rooij

**Coherent Multimodal Couplings in the Urban Web for Successful Cities**
R.M. Rooij

**Life Cycle Management Systems**
A. Well Balanced Methodology to Navigate in Complex LOGISTIC Waters
J. Stavenhuisen

**Visual Search Difficulties as Causative Factor of Intersection Accidents in Rice Fields**
N. Uchida

**Driver Support Systems, Regulation and Policy**
K.A.P.C. van Wees

### Part 2

**Operating Models for Dedicated Transportation Systems and Their Implications**
A.J. van Binsbergen
J.G.S.N. Visser

**Implementation Strategies for New-generation Inland Terminals**
Y.M. Bontekoning

**Competitive River-sea Transport**
J.W. Konings
M.W. Ludema

**Innovative Networks and New-generation Terminals for Intermodal Transport**
E. Kreutzberger

**TRACES**
The Agile Traffic-control and Engineering System
J.J.M. Evers
D.G. Lindeijer

**Integrated Environmental impact Modelling for Urban Freight Transport**
L. Ma

**Flexible and Robust Control of Logistic Processes**
J.M. Valk

**The Design Of Logistic Control for Intermodal Transport Chains of the 21st Century**
C. Versteeght

**Towards Algorithms for Scheduling of Transport Processes**
M.M. de Weerdt

**The Service-Oriented Agile Logistic Control and Engineering System: SERVICES**
J.J.M. Evers
L. Loeve
D.G. Lindeijer

**Analysing Logistical Trends in the Dutch Food Industry**
G.J. Muilerman

**Electronic Auctions, Market Efficiency and Strategic Impact**
O.R. Koppius

**Virtuality and Inter-Organisational Co-operation in Logistics and Transport**
T.M. Verdulijn

**The Business of Modularity and the Modularity of Business**
M.J.J. Wolters

---

Parts 1 & 3 of these Conference Proceedings are published in separate volumes.
Preface

By Professor Piet H.L. Bovy
Scientific director
TRAIL Research School

The TRAIL Research School for Transport, Infrastructure and Logistics, the co-operative School for Ph.D. research and education of Delft University of Technology, Erasmus University Rotterdam and the University of Groningen, carries out fundamental research in a diversity of areas organized in a number of coherent programmes. Each year, the progress achieved in these programmes is presented at the yearly TRAIL Congress by written papers and oral presentations of the Ph.D. students aimed at informing a wider audience of scientists and professionals. Submitted contributions are thoroughly screened, selected, and improved with the help of an International Review Committee. The contributions included in this volume constitute the outcome of this careful process (see also Parts 2 and 3 of these Proceedings issued in separate volumes).

The reported research is centered around a number of joint research programmes in which the researchers co-operate. In the multi-disciplinary programme on Automatic Vehicle Guidance (AVG) interesting contributions are given with respect to roadway capacity impacts of AVG, product liability issues of AVG, and on technology assessment of this emerging technology. The related research programme on Freight Transport Automation and Multimodality (FTAM) is represented by a series of papers covering subjects such as the design and performance of new-generation intermodal freight terminals, design of new load units and logistic concepts for short distance intermodal freight transport services. Another larger TRAIL programme concerns Seamless Multimodal Mobility (SMM) of personal travel. At the strategic level, related contributions deal with travel choice modelling, optimizing mixed public transport service networks, and design of multimodal transfer points. At the operational level of SMM, papers deal with delay modelling of trains in railway networks, with establishment of delay-robust time tables, and with the establishment of decision support systems for synchronization control of scheduled train services.

In the area of road traffic flow analysis and control a number of papers tackle the important but difficult subject of modelling multiple user class traffic flow operations leading to extended theories of traffic flow. The policy analysis research stream of TRAIL is represented by a variety of contributions related to infrastructure planning and transportation decision making in the public and private domains. These partly deal with methodological issues of policy analysis and partly with substantial issues such as the consideration of safety and risk aspects in multimodal corridors, or urban development in high-speed train station areas. A final larger programme of TRAIL research represented in these proceedings with several contributions is on Logistic Systems Control. An interesting example of this deals with time strategies in logistics.

The collection of papers is an impressive demonstration of the rapid and valuable progress made by the collective efforts of the Ph.D. researchers. Their results constitute the bricks from which the building of knowledgement will be constructed to the benefit of science and society.
The Netherlands TRAIL Research School, 1999
International Review Committee (IRC)

The review
All papers submitted by TRAIL Ph.D. students to the TRAIL yearly Congress are subject to a review procedure. The review process aims at improving the quality of the submitted papers (and related research), contributes to the selection of papers for the Congress, and results into nominations of papers for the T&L Awards.

The best papers are published by the Delft University Press in their TRAIL Conference Proceedings Series (Part 1 of these Congress Proceedings) and are recommended for publication in international scientific journals. The remaining accepted papers are published in separate volumes of the Congress Proceedings.

The review process is carried out under the responsibility of the scientific director of TRAIL Research School (currently Professor Piet H.L. Bovy) and the Managing Director of TRAIL, André L. Loos, with the assistance of the TRAIL Programme Board and with the invaluable help and efforts of the TRAIL International Review Committee (IRC).

In 1999 the following members of this committee contributed to the paper reviews.

Dr. ir. B. van Arem (TNO-Inro, The Netherlands)
Prof. dr.-ing. K.W. Axhausen (IVT Zurich, Switzerland)
Ir. R. Babuska (Technische Universiteit Delft, The Netherlands)
Prof. D. Bannister (University College London, United Kingdom)
Dr. M.G.H. Bell (University of Newcastle, United Kingdom)
Prof. dr. ir. J.W.M. Bertrand (BETA, The Netherlands)
Prof. dr. M. Beuthe (FUCAM, Belgium)
Dr. V.J. Blue (New York State University, USA)
Prof. dr.-ing. W. Brilon (Ruhr-Universität Bochum, Germany)
Prof. A. Ceder (Technion – Israel Institute of Technology, Israel)
A. Daly (Hague Consulting Group B.V., The Netherlands)
Dr. L. Dam Hanh (Institute for Transport Policy Studies, Japan)
Dr. M.J. Dijkstra (NETHER, The Netherlands)
C.L. Felbinger (The American University Washington, USA)
Dr. ir. Filarski (Adviesdienst Verkeer en Vervoer, The Netherlands)
Prof. N. Gartner (University of Massachusetts, USA)
Dr. S.A. Hakkert (Technion – Israel Institute of Technology, Israel)
Prof. dr. ir. R. Hamerslag (The Netherlands)
M. Hasebe (Institute for Transport Policy Studies, Japan)
Prof. R. Hassin (Tel Aviv University, Israel)
Dr. ir. G. van den Heuvel (Nederlandse Spoorwegen N.V., The Netherlands)
Prof. B.G. Heydecker (University College London, United Kingdom)
Dr. C.P.M. van Hoesel (Universiteit van Maastricht, The Netherlands)
Prof. dr. A.I.J.M. van der Hoorn (Universiteit van Amsterdam, The Netherlands)
Ir. W. Janssen (TNO Technische Menskunde, The Netherlands)
Mr. J.V. Jones (Logistic Management Associates, USA)
Prof. Kap Hwan Kim (Pusan National University, Korea)
Dr. C. Kara-Zaitri (University of Bradford, United Kingdom)
Dr. ir. E.A.F. Kraan (Andersen Consulting, The Netherlands)
Drs. B. Kuipers (Technische Universiteit Delft, The Netherlands)
TRAIL management and TRAIL Programme Board are grateful to these members of the International Review Committee for their efforts and invaluable comments and suggestions for improving the submitted papers.
The Free Speed Distribution of Drivers: Estimation Approaches

TRAIL Research School, Delft, December 1999

Author
Ir. H. Botma
Faculty of Civil Engineering and Geosciences, Delft University of Technology

Thesis supervisor
Prof.dr.ir. P.H.L. Bovy
Faculty of Civil Engineering and Geosciences, Delft University of Technology
The Flexi Speed Distribution Grid

The Paper is sponsored by the TRAIL Research School Policy Development Group.

TRAIL management and TRAIL Programme Board are grateful to those members of the International Review Committee for their efforts and invaluable comments and suggestions underlying the submitted papers.

Author:

Title:

In charge of Commissioning and Operation

Title of the Paper:

I, T. E-mail

Title:

Author:

Date of technical report:

Prepared by:

Title:

I, T. E-mail

The Netherlands TRAIL Programme Network May
Contents

Abstract

1 Introduction 1

2 Concept of free speed 2

3 Relevance of free speeds 3
  3.1 Traffic flow models 3
  3.2 Design 4

4 Free speed choice of individual drivers 5

5 Empirical estimation methods of the free speed distribution 8
  5.1 Speeds at Low Intensities 8
  5.2 Extrapolation Towards Low Intensities 9
  5.3 Method Based on a Simulation Model 9
  5.4 Method Based on Erlander's Model 9
  5.5 Method Using Censored Observations 10

6 Earlier investigation of the censored estimation method 12

7 Criteria for separating free and constrained vehicles 13

8 Applications of the censored method 15

9 Conclusions and discussion 19
  9.1 Conclusions 19
  9.2 Further studies 19

10 References 21
Abstract

Free speeds are defined as the speeds at which vehicles are driven when the drivers are not influenced by other vehicles. As driver-vehicle combinations have different characteristics that influence their choice of free speeds, the outcomes will be different and the free speed can be described as a stochastic variable with a distribution. This distribution plays an important role in many traffic flow models, macroscopic as well as microscopic ones.

A utility model that describes the free speed choice of drivers is discussed. However, a free speed distribution cannot be derived from this type of models and more direct empirical methods are needed.

In general only a part of the drivers in a traffic flow will be able to drive at their free speed. The speeds of the vehicles that drive free are not representative of the free speed distribution of all drivers on the road because the lower free speeds are over-represented.

The paper discusses several methods to estimate the free speed distribution and works out the method based on censored observations. This method has been evaluated by using two types of simulation models and has been applied to data from high volume two-lane rural roads.

The estimation method requires the vehicles to be separated in two classes: free driving and constrained driving. Several criteria to do this are discussed and two extremes are compared regarding their influence on the outcome of the estimation.

Keywords: traffic operation, free speed distribution, estimation, censored observations
1 Introduction

This paper will discuss three items.

1. Why does a 'lonely vehicle', i.e. a vehicle not influenced by other road users, on a given road drive as fast as it does? In other words, how does a driver choose his/her free speed? Can we understand this choice process, relate the outcome to observable variables, and derive an operational useful model? This problem is especially relevant when road authorities consider speeds driven to be too high and want to change the road characteristics such that speeds become lower. An alternative measure to control speeds is of course a speed limit with enforcement but the latter will not be considered here.

2. Different driver-vehicle-combinations (DVC's) appear to have different free speeds and this is practically described by a free speed distribution, to be denoted by FSD. Empirical methods to determine this distribution are faced with the difficulty that in practical situations quite a few DVC's are not driving at their free speed because other vehicles prevent that. The simple way out is to collect only speeds of free driving vehicles. However, this is not always correct and can lead to substantial errors. The speeds of free driving vehicles are not a representative sample of the unknown FSD but a sample with a distortion, because a DVC with a high free speed is less likely to be observed in a free state than a DVC with a relatively low free speed. The way to overcome this distortion is the main subject to be discussed in this paper.

3. Dividing vehicles passing a cross-section in free driving and constrained driving is not obvious. Several criteria have been proposed and some of them will be discussed. In this context it is relevant if the criterion used has much influence on the outcome of the estimation of the FSD.

From a practical point of view most interest in FSD is at motorways and other main arterials for motorised traffic only. We will limit the practical discussion in this paper to two situations: a carriageway of a motorway, hence at least two lanes for one-directional traffic, and two-lane rural roads, hence two lanes for two-directional traffic.

In section 2 of this paper we will discuss the concept of free speed and in section 3 its practical relevance for traffic flow models and road design. Next a survey of methods to estimate the FSD is presented and the method based on censored observations is worked out. Section 7 discusses criteria to separate free driving vehicles from constrained vehicles. The censored method is applied to data from two-lane roads. The paper ends with some preliminary conclusions and sums up further required research.
2 Concept of free speed

Usually the 'free speed' or 'desired speed' of a driver-vehicle combination (DVC) is defined as the speed driven when the driver is not influenced by other road users in the choice of his or her speed. We will use the term free speed in this paper and reserve the term 'desired speed' for a different parameter.

The free speed driven will be influenced by properties of the vehicle, the driver, the road, conditions such as weather and rules such as speed limits.

Sometimes a more comprehensive definition is used: free speed is the speed driven when there is no influence of other road users and no influence of special local road characteristics such as a sharp curve, a steep grade, a narrow bridge etc.; see for example Gynnerstedt (1976). This difference of definition is useful but has little relevance for this paper.

According to both definitions the free speeds on a freeway will generally be higher than on a two-lane road, which can be ascribed to the general and permanent differences between these two types of road.
3 Relevance of free speeds

Knowledge of free speeds at a road under given conditions can be relevant for several reasons; we will discuss here traffic flow models and design.

3.1 Traffic flow models

The concept of free speed appears to be an important theoretical notion in current traffic flow theory. Consequently the free speed distribution is an important element for many traffic flow models, analytical ones as well as simulation models. We will give a short overview.

3.1.1 Microscopic simulation models

Usually every vehicle is assigned several individual characteristics at its entrance in the simulation process of which free speed is one. If all free speeds are the same, the model will not behave realistically because no incentive to carry out overtakings or lane changes is present. A driver is assumed to try to maintain his/her free speed as long as possible and will accelerate to it after an inevitable speed reduction. An example is the model developed by Minderhoud (1999) to investigate the effect of intelligent cruise control on traffic flow behaviour.

Mostly, free speeds are modeled as drawings from a Gaussian distribution with parameters dependent on vehicle type and driver type. The idea that the free speed of a DVC is correlated with other assigned behaviour parameters is attractive but empirical knowledge about the relation between free speed and e.g. characteristics of the car-following behaviour is hardly present.

3.1.1.2 Mesoscopic models

The distribution of free speeds is a key element in the mesoscopic traffic flow models of which the development more or less started with the work of Prigogine & Herman (1971). The models describe how the speed distribution is changing from the FSD at intensity zero to a distribution at intensities at which vehicles are only temporarily driving free. A main problem with these types of models is that in dynamic situations the speed distribution loses its operational applicability, i.e. the size of the time-space areas that are relevant for the traffic flow behaviour can be so small that they contain only a few vehicles. This is an important difference with Boltzmann type models describing phenomena in gasses.

However, recently developed mesoscopic models have shown to be a good starting point for developing more comprehensive macroscopic models, as shown by Helbing (1996) and Hoogendoorn (1999).

1) In stead of flow, rate of flow, or volume the term intensity will be used for the number of vehicles passing a cross-section per time period.
3.1.1.3 Macroscopic models

The mean free speed is a parameter of the most basic relation in macroscopic traffic flow theory, the fundamental diagram. The latter term has been introduced by Haight (1963) who derives a form of this diagram by assuming that free speeds have a gamma distribution that is being modified in a logical way as density increases from zero to jam density.

In macroscopic models derived from mesoscopic relations, the free speed distribution or its parameters play a direct or indirect role; examples are the macroscopic dynamic model of Helbing (1996) and the multilane multiclass dynamic model of Hoogendoorn (1999).

3.2 Design

In the design process for a new road, the function of the road in the network and the predicted design intensity determines the suitable type of road. Every element of the road should be designed in such a way that vehicles can drive along safely and comfortably at the design speed, which is usually a percentile of the free speed distribution. In fact, it should be verified afterwards if the FSD of the road in use equals the assumed FSD. The author has the impression that this point gets little attention in practice.

Often speeds are considered too high for traffic safety reasons. The relation between the speed limit chosen and the FSD of the road is of importance. If e.g. the speed limit equals the 15 percentile of the FSD, it can be predicted that the level of enforcement has to be high, otherwise the limit will hardly have any impact.
4 Free speed choice of individual drivers

To introduce behavioural models that may explain and predict free speeds, we introduce the notion of free speed again. Consider the speed of a Driver-Vehicle-Combination (DVC) at a cross-section $x$ of a road at time $t$: $v_i(x,t)$. Factors that influence this speed can be broadly classified into:

a. Characteristics of the DVC
   - Permanent driver characteristics: experience, temperament, knowledge of the road, etc.
   - Temporary driver characteristics: trip purpose, mood, level of being in a hurry, etc.
   - Vehicle: type, within type power, maximum speed, etc.

b. Road characteristics
   - Type of road; within type: road width, horizontal and vertical profile, quality of surface, etc.

c. Other traffic
   - Vehicles in front that cannot be overtaken;
   - Vehicles in adjacent lanes, either opposing or driving in the same direction;
   - Crossing traffic.

d. Traffic rules
   - Most important in this context is the speed limit.

Consider a situation in which factors of class a, b and d are practically constant and think away the other traffic. Also assume that the road characteristics are rather constant along the road. In other words we consider the speed of a lonely DVC at a homogeneous section of a road; to limit the situation even further we assume a road only meant for use by fast traffic (cars and trucks) and raise the point: why does the driver go as fast as he/she does?

We will define this speed as being the free speed of the driver. This term is preferred over desired speed because one can argue that the real desired speed is infinite. Several models have been developed with the aim to (quantitatively) explain and predict the free speed of drivers. A relatively old one and a recent one will be briefly discussed.

Volmuller (1976) has proposed a model that explains the free speed of a driver as a result of being subjected to a number of counteracting mental stresses:

1. Stress caused by the desire to reach the destination. The higher the speed the lower the stress; see Figure 1. Of course this curve can be on different levels and certainly is dependent on the type of trip and the time schedule of the driver. This stress is zero at infinite speed, the ultimate desired speed.
2. Stress due to physical causes, e.g. noise, vibrations, vertical and lateral accelerations and jerks. It is evident that this stress is very dependent on the quality of the road and the vehicle and that it increases with speed, at least from a certain speed level.

3. Stress due to the driving task. The higher the speed the more requiring is the task of perceiving all cues in time to stay on the road, respond in time to the sudden appearance of other road users (e.g. crossing deer). Again the form of this stress curve will depend on characteristics of the DVC and especially the knowledge of the driver of the road characteristics.

Consequently Figure 1 presents only possible curves but as a rule they sum up to an overall stress level that has a minimum at the free speed. This model can be of use for road designers that consider measures to reduce speeds in a 'natural' way without using such drastic measures as road humps, etc. Jepsen (1998) has proposed a model of the same conceptual type. He introduces more factors, e.g. the probability of being snapped by the police and the expected fee; the probability of a car breakdown; the probability of an accident, etc. The practical result is a model of the fundamental diagram of traffic on a motorway with some 20 parameters, which does not make it a very efficient tool. But the usefulness of this model lies in its potential explaining power of predicting the direction a measure can have on speeds and possibly also which measures have more effect than others.
Especially the role of the speed limit and the enforcement is probably a main factor. Nowadays cars on modern roads allow high speeds, say 150 km/h and more, without causing much discomfort for the driver and the sensation of high speed seems to be appreciated by many drivers. Consequently it becomes very important how drivers respond to the speed limit and its enforcement.

Studies to transform these conceptual models into a practical quantitative tool have been carried out but never have been very successful. A lot of practical information is available about the level of speed in given conditions but this mostly can be typified as empirical relations. According to the opinion of the author it will be virtually impossible to obtain a free speed distribution using this type of models. They are a help in understanding qualitatively the process of free speed choice but can not be made sufficiently operational.

2) I.e. a relation between e.g. road characteristics and the speed of drivers without modelling the driver perception, processing and response to the cues as an intermediate step.
5 Empirical estimation methods of the free speed distribution

Since different drivers exhibit different free speeds individual values are not of much use in applications. The FSD appears to be an adequate description. Given an operational criterion (see section 6 for a discussion) for driving free at a cross-section, vehicles passing by can be separated in two categories:
- vehicles driven at their free speed (so-called free vehicles or free drivers);
- driver-vehicle combinations that are car-following (so-called constrained vehicles).

It should be remarked that this separation in two groups is an approximation of reality; there are several types of states between free and constrained feasible.
The observed distribution of the speeds of free drivers would be a good estimator for the FSD, if these free drivers were a random sample of the driver population at the road considered. However, this is not true because a driver with a relatively high free speed has more chance to become constrained than a driver with a relatively low free speed. Stated in other words, platoon leaders have a relatively low free speed and the followers in the platoon have a relatively high free speed, at least it is higher that the free speed of the platoon leader. This is the essential problem in determining the FSD and several proposed solutions for this problem will be discussed.

5.1 Speeds at Low Intensities

At low intensities most drivers will be able to drive their free speed. The distortion mentioned earlier, that vehicles with a high free speed are more often constrained, is not of much importance in that case. Consequently at low intensities the observed distribution of free speeds is nearly equal to the FSD.

In order to have a high fraction of free drivers the intensity has to be low. What is sufficiently small depends on the type of road.
At motorways the presence of two or more lanes for one direction allows drivers to drive most of the time free. According to the HCM, TRB (1994), this can be assumed to be the case for intensities up to 1400 veh/h per lane. Hoogendoorn (1999) is more cautious and applies a limit of 400 veh/h per lane. Using the HCM limit offers plenty of opportunities to collect sufficiently large samples at the right representative periods.

In contrast, at two-lane roads the fraction of constrained vehicles increases steeply with intensity. According to the Highway Capacity Manual of 1985 and Dutch results, Botma (1986), at an intensity of 100 veh/h per lane already approximately 15% of the vehicles is constrained. This fact implies that on many two-lane roads the method can only be used in the evening or at night.

---

3) Strictly speaking it would be better to use low densities but if congested flow is excluded low intensities refers roughly to the same state of traffic flow.
At low intensities the population of drivers using the road can be quite different, for example regarding trip purpose, from the population at more busy hours. It does not seem justified to assume that the FSD is the same during all hours of a day.

Generally traffic flow models are most relevant at peak hours with busy traffic or even congestion. The values of important parameters of these models, of which the FSD might be one, should be representative for the periods and conditions of interest and it is questionable whether they are the same for the peak-driver and the non-peak driver population.

5.2 Extrapolation Towards Low Intensities

It is possible to establish relations between the intensity and the speed distribution for periods with relatively high intensities that are of interest. By means of extrapolation to low intensities, in fact to intensity zero, the FSD can be obtained.

The advantage of this method, compared to the former, is that now data from the relevant driver population are used. The disadvantage of the method is that relations between e.g. mean and standard deviation of speed on the one hand and intensity at the other, are characterised by large fluctuations and consequently an extrapolation will not be precise. This disadvantage is especially relevant at two-lane roads.

5.3 Method Based on a Simulation Model

Hoban (1980) used the Australian microscopic simulation model TRARR for traffic operation on two-lane roads, to establish relations between observable variables, such as intensity and mean speed, and the FSD used as an input. Using these relations in the opposite direction the FSD can be calculated. This method seems feasible, also for other road types, but depends on the validity of the model used.

5.4 Method Based on Erlander's Model

Erlander (1971) developed an integral equation for traffic operation at two-lane roads with the FSD as a component. The model is rather schematic and relations with the FSD involve parameters of overtaking behaviour and platooning behaviour that are difficult to obtain in field studies. McLean (1989) concludes the method can not yet be used in practice.

To conclude: all methods discussed so far have disadvantages. These are especially relevant at two-lane roads. At motorways the first method, using data at low intensities, might be sufficiently valid for practical use. However, it is worthwhile to compare the outcomes of this simple approach with an application of the next method.
5.5 Method Using Censored Observations

Recall that we want to estimate the free speed distribution; we will denote its probability density function (PDF) by \( f(v) \) and its distribution function by \( F(v) \).

Suppose one has observed speeds of vehicles passing a cross-section together with the state of the vehicle, i.e. free or constrained. Speed can be measured directly while state can be deduced simply from time headway or more complex by considering also relative speed of the vehicle w.r.t. the vehicle in front and the possibilities to overtake or change lane; see section 7.

A speed of a free vehicle \( v_i \) can be considered as a drawing from the unknown PDF \( f(v) \) and contributes a term \( f(v_i) \) to the likelihood of the sample.

A constrained vehicle has a speed \( v_i \) that is, by definition, less than its free speed. If we knew how much there would not be a problem. Its likelihood is the probability that a free speed is larger than \( v_i \), consequently its contribution to the likelihood of the sample is: \( 1 - F(v_i) \).

If we sort the sample to \( i = 1, 2, ..., n_1 \) free speeds followed by \( i = n_1 + 1, ..., n \) constrained speeds, the total likelihood of the sample is:

\[
L(v; \theta) = \prod_{i=1}^{n_1} f(v_i; \theta) \prod_{i=n_1+1}^{n} (1 - F(v_i; \theta))
\]

Assuming e.g. a Gaussian distribution with mean \( \mu \) and standard deviation \( \sigma \), functions \( f \) and \( F \) are known and parameter \( \theta = (\mu, \sigma) \) can be calculated by some method to maximize the likelihood function \( L \), or more simply, maximize its logarithm.

The procedure given is a straightforward application of the estimation theory of censored observations; see e.g. Nelson (1982). The observed free speeds are ordinary or uncensored observations, the speeds of constrained vehicles are so-called censored observations.

Note 1:

The censored estimation method can also be used without specifying the mathematical distribution function; then it is called Product Limit Estimation method. It turns out that when the sample is not too small the outcomes of both approaches are not that much different; see Botma (2000).

Note 2:

Branston (1979) derived a method to estimate the FSD, starting from a result from Cowan's (1971) model for single lane traffic adding an ingenious interpretation. It appears to be exactly the same estimation method as the PLM; see Botma (2000).
Note 3:
The same estimation procedure can be used in two other traffic analysis problems:

1. Estimation of the capacity distribution of a bottle-neck from measured capacities and measured high intensities; see Minderhoud et al (1998). The observed capacities are the ordinary observations whereas the high intensities (we only know that the capacity at that period is higher) are the censored observations.

2. Estimation of the critical gap distribution. A rejected gap is by definition smaller than the critical gap and an accepted gap is by definition larger. So in this case we have only censored observations but because they are censored to the left and to the right they can be used to estimate the unknown distribution. This method has already been proposed by Miller (1972) and has been recently rediscovered, Brilon et al (1997), and proven to be superior to many well known classical methods.
6 Earlier investigation of the censored estimation method

Earlier investigations (Botma, 1987) have concentrated on Branston’s derivation of the estimation method which only later was found to be the same as the formal statistical PLM method. The tool used was a simple simulation model describing the movement of vehicles over a single-lane road section without overtakings. Using this model several assumptions of Branston have been investigated and we repeat the main results:

- In Branston’s derivation, section speeds were replaced by local speeds at a cross-section. This leads to a small but systematic error in the estimated FSO.
- It can be assumed that each driver has its own limits between free and constrained driving. A second assumption of Branston is these individual limits can be replaced by a collective one and this conjecture proved to be right.
- The method assumes that for vehicles at the entrance of the road section without overtakings no correlation between free speed and state exists. Introducing such a correlation in an ad-hoc way did not seem to impair the estimation method.

As a second tool to assess the estimation method, the microscopic simulation model TRARR mentioned earlier, is in use in the current study. A preliminary result of this work in progress is that the estimation method shows a positive bias, i.e. the mean as well as the standard deviation of the FSD are somewhat overestimated. It has been possible to reduce the bias in the mean to a negligible size by giving the observations of speed of constrained vehicles in platoons a weight that decreases with the platoon size. However, in doing so the bias of the standard deviation increased.

Note: The fact that estimating the FSD is a formal statistical method does not imply it is correct or of sufficient quality. The usual assumption that the sample should consist of independent elements is violated in this case; speeds in a platoon are certainly highly correlated. And the relation between the state of the vehicle and its free speed is very much hidden in the traffic flow process we want to understand and model. So there might be a type of vicious circle here, especially when using a simulation model to investigate the estimation of the FSD. On the other hand the FSD of the driver population at the representative conditions is and remains unobservable, consequently simulation models are an essential tool.
Criteria for separating free and constrained vehicles

We will present a short overview of practical methods to divide vehicles observed at a cross-section in free and constrained. It is assumed that speed and time headway are observed.

7.1.1.1 Time headway

The idea to use time headway as a criterion is already present in the HCM of 1950 in which a nowadays surprisingly high value of 9 s is mentioned. In the HCM of 1985 the criterion plays an important role in analysing traffic operation at two-lane roads for which a value of 5 s has been chosen. In the new HCM of 2000 this value might be decreased to 3 s. Dutch guidelines, Commissie RONA (1992), use also 5 s. A background study, Botma (1986), showed it is difficult to pick a value based on its consequences because relations at 4, 5 and 6 s were 'congruent'. Dijker et al (1998) used different limits on motorways for cars (3.5 s) and trucks (5 s). Finally it should be realised that this criterion loses its validity at congestion.

7.1.1.2 Time headway and relative speed

The notion that the relative speed should not be large in the constrained state was made operational in an OECD (1972) study. The general idea is that large relative speeds go together with lane changings or overtakings.

7.1.1.3 Fuzzy interpretation of margin of required and actual distance headway

Jepsen (1998) defines a safe or required distance headway as the sum of: the vehicle length $L_{veh}$; a constant distance $s_0$ at speed $v$; the distance covered during the response time $r$; and the parameter "speed risk factor" $\alpha$' times the squared speed. The actual distance headway equals speed $v$ times time headway $h$. Consequently the margin equals:

$$m = s_{\text{actual}} - s_{\text{required}} = vh - (L_{veh} + s_0 + rv + \alpha v^2)$$

Hoogendoorn (1999) uses this margin of Jepsen (with $\alpha = 0$) and adds the relative speed as a second criterion. Both criteria are transformed into a 'level of constrainedness' by means of two membership functions. They are defined in such a way that a vehicle is more free if its distance margin and its relative speed are larger.
The new point in this approach is that a vehicle can be partly constrained. Fortunately the censored estimation method of the FSO can handle partly constrained vehicles without modifications. Denote the level of constrainedness or 'state' by $S$; $S=1$ for a totally constrained vehicle and $S=0$ for a totally free vehicle. Then a different way of writing eq.(1) allows state $S$ to be a fraction:

$$L(v, S; \theta) = \prod_{i=1}^{n} f(v; \theta)^{1-S} (1-F(v; \theta))^S$$

(3)

7.1.1.4 Probability contours in the time headway-relative speed plane

Botma et al (1980) used detailed vehicle data over a 1200 m long road section to determine, based on graphs of speed, headway and relative speed as a function of position, if the vehicle considered was free at the cross-section in the middle. This results in a probability of being constrained using headway and relative speed of one cross-section; see Figure 2. So in fact the type of result of this approach and the fuzzy procedure are similar.

Figure 2: Contours of the probability function giving the probability that a vehicle is constrained based on its time headway and relative speed w.r.t the vehicle in front.
8 Applications of the censored method

Research is in progress of applying the censored estimation method, using data from two-lane rural roads and motorways. Some preliminary results of two sites at two-lane rural roads, flat and straight road sections with characteristics according to design standards, will be presented. Some general characteristics of the sites are presented in table 1. Site I and site II reflect measurements at off-peak hours with relatively low intensities and measurements at peak hours with relatively high intensities respectively. Both sites have about the same truck percentage.

<table>
<thead>
<tr>
<th></th>
<th>Lane Intensity</th>
<th>Sample Size</th>
<th>Opposing Intensity</th>
<th>Truck Percentage</th>
<th>Mean Speed</th>
<th>St. Dev. Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site I</td>
<td>veh/h</td>
<td>veh</td>
<td>veh/h</td>
<td>%</td>
<td>km/h</td>
<td>km/h</td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>660</td>
<td>300</td>
<td>17.2</td>
<td>97.2</td>
<td>14.3</td>
</tr>
<tr>
<td>Site II</td>
<td>790</td>
<td>1931</td>
<td>920</td>
<td>18.9</td>
<td>84.8</td>
<td>11.0</td>
</tr>
</tbody>
</table>

Main results are best presented by a graph of several distribution functions; see Figures 3 and 4. From left to right are depicted:
- speed distribution of constrained vehicles;
- speed distribution of all vehicles;
- speed distribution of free vehicles;
- estimated FSD.

The criterion applied to separate constrained from free vehicles in this case is a time headway of 4 s. It can be seen that the higher the intensity, the larger the difference between speeds driven and the FSD. Table 2 presents the numerical estimation results and the standard errors obtained with Bootstrap. The standard error of the mean of the FSD seems quite acceptable; the standard error of the estimated standard deviation of the FSD is 4 to 5 times higher but also still acceptable.
Table 2: FSD Estimation Results; Criterion is Time Headway of 4 s

<table>
<thead>
<tr>
<th>Site</th>
<th>Fraction Constrained (%)</th>
<th>Mean of FSD (Km/h)</th>
<th>St. Dev. of FSD (km/h)</th>
<th>St. Error of Mean (%)</th>
<th>St. Error of St. Dev. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site I</td>
<td>38.0</td>
<td>104.1</td>
<td>15.9</td>
<td>0.8</td>
<td>3.6</td>
</tr>
<tr>
<td>Site II</td>
<td>75.5</td>
<td>100.9</td>
<td>13.9</td>
<td>0.6</td>
<td>3.2</td>
</tr>
</tbody>
</table>

Figure 3: Speed distributions at low intensity site; from left to right: speed distribution of constrained, all, and free vehicles, and estimated FSD
The estimation method can easily be applied to observations of cars and trucks separately. It can be expected that trucks have lower free speeds than cars and consequently are more likely to be free driving. Table 3 presents the results and it can be seen that they are according to these general expectations. In fact there is not that much difference between the results of the two sites. It is remarkable that the mean free truck speed is not so far below the mean free car speed. It can be added that the speed limit at Site I is 100 km/h and at Site 2 80 km/h; this difference in speed limit seems to have no effect on the FSD.

Table 3: FSD estimation results for cars and trucks; criterion is time headway of 4s

<table>
<thead>
<tr>
<th>Site</th>
<th>C</th>
<th>Mean of FSD</th>
<th>St. Dev. of FSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cars</td>
<td>48</td>
<td>106</td>
<td>17</td>
</tr>
<tr>
<td>Trucks</td>
<td>24</td>
<td>97</td>
<td>13</td>
</tr>
<tr>
<td>Site II</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cars</td>
<td>79</td>
<td>104</td>
<td>17</td>
</tr>
<tr>
<td>Trucks</td>
<td>60</td>
<td>96</td>
<td>15</td>
</tr>
</tbody>
</table>
Table 4: Effect of Headway and Fuzzy Criterion on FSD Estimates

<table>
<thead>
<tr>
<th>Site</th>
<th>Fr. Constr. %</th>
<th>Mean FSD km/h</th>
<th>St. Dev. FSD km/h</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>33.0</td>
<td>103.0</td>
<td>15.5</td>
</tr>
<tr>
<td></td>
<td>38.0</td>
<td>104.0</td>
<td>15.9</td>
</tr>
<tr>
<td></td>
<td>43.0</td>
<td>105.0</td>
<td>16.4</td>
</tr>
<tr>
<td></td>
<td>32.0</td>
<td>103.0</td>
<td>15.2</td>
</tr>
<tr>
<td>II</td>
<td>68.0</td>
<td>98.0</td>
<td>13.4</td>
</tr>
<tr>
<td></td>
<td>76.0</td>
<td>101.0</td>
<td>13.9</td>
</tr>
<tr>
<td></td>
<td>80.0</td>
<td>103.0</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td>73.0</td>
<td>100.0</td>
<td>13.0</td>
</tr>
</tbody>
</table>

Finally Table 4 presents the effect of the constrained/free criterion. The larger the time headway criterion, the larger the fraction constrained vehicles, the estimated mean free speed and the estimated standard deviation. But the effect of the criterion on the fraction constrained vehicles is larger than on the mean and st. dev. of the FSD, which is a useful property. Results of application of the fuzzy criterion do depend on the values of its 8 parameters. They have been chosen based on values used by Jepsen (1998) and Hoogendoorn (1999) and modified in such a way that the fraction of constrained vehicles was inside the range produced by the time headway criterion. It can be concluded that in this case the results of the application of the fuzzy criterion do not result in different estimation results. The preliminary conclusion can be that the simple criterion is suitable.
9 Conclusions and discussion

The concept of free speed is an important notion in traffic flow theory. However, empirical estimation of this variable is not self-evident, especially because the free speed distribution is an unobservable parameter that precludes a simple comparison of several methods against one true reference. Several estimation approaches have been reviewed and discussed.

9.1 Conclusions

Given that the study into estimation methods of the FSD is still in progress, the conclusions will have a preliminary character.

• Among the methods considered to estimate the FSD the one based on the concept of censored observations seems to be the best founded.

• The criterion to separate constrained and free vehicles has only a minor influence on the outcome of the estimation. So it seems a simple criterion is suitable.

• It is possible to estimate the FSD's for separate vehicle categories.

• Estimation errors can best be estimated using Bootstrap because classical methods to estimate standard errors depend too much on assumptions that are not valid in this application.

9.2 Further studies

At a motorway the state of a vehicle also depends on characteristics of vehicles at nearby lanes to which a constrained vehicle could make a lane change. E.g. a vehicle at the slow lane closely following a vehicle in front is definitely constrained according the criteria discussed before. But if this vehicle has no problem to change to the faster lane because the gap there is large enough, would it be correct to classify it as constrained? Consequently the criterion to separate free from constrained vehicles should be extended, especially at motorways taking into account the lane change possibilities of the vehicle to be classified. It should be an operational criterion that preferably can be applied using data from ordinary detector stations equipped with induction loops.

For motorway traffic the simple method to estimate the FSD, i.e. assuming the free driving vehicles are the right sample, should be compared with the outcomes of the censored estimation method.

A microscopic simulation model for motorways should be used to investigate the different estimation methods using the same procedure as is in progress for two-lane roads. The model should have sufficient detail regarding generation of the free speeds and the assignment of the free speeds to the vehicles.

Finally some general discussion points. i.e. questions and no answers.
9.2.1.1 How important is it to have the right FSD?
This will depend on the application and could be different for design and modelling purposes. In fact this question can be put forward for every parameter in a traffic flow model. The first step to gain knowledge is a sensitivity analysis.

9.2.1.2 Applicability of the concept of free speed
In fact the concept of a free speed as a driver-vehicle attribute is debatable. How constant is a free speed of a DVC. If it is variable and depends too much on traffic conditions, then its usefulness as a parameter in models can be questioned. The author's expectation is that the more the concept of free speed and FSD is analysed and detailed, the less its applicability for practical traffic engineering will become. On the other hand: a response time of a DVC is a comparable parameter and an element of many traffic flow models. Its relation with the real response time of different DVC's, being exposed to different traffic experiences during the last period of, say, 2 minutes, is as debatable as the value of a free speed.

9.2.1.3 Will free speed be a relevant variable in the future?
There is a growing tendency of controlling speeds of individual vehicles. The technical possibilities do exist to control the actual maximum speed of each vehicle and make it dependent on the road characteristics, the traffic conditions, the weather, etc. In principle the total traffic system could work better in that case but the drivers freedom would be substantially impaired. It can be expected that the car industry does not favour this idea.
10 References


• Commissie RONA. Guidelines for the design of rural roads; motorways excluded. SDU, Den Haag, 1992 (in Dutch).


• Erlander, S. A mathematical model for traffic on a two-lane road with some empirical results - II. Empirical results. Transportation Research 5(1971)2:149-175.


• OECD. Speed limits outside built-up areas. OECD, Paris, 1972.


Abstract

1 Modelling and estimating dynamic origin-destination matrices

1.1 Introduction

1.2 The scope for further work in O/D matrix estimation

1.2.1 Relevance of transportation modelling

1.2.2 Expected benefits from further work in O/D matrix estimation

2 Using models to describe object systems

Modelling and Estimation of Dynamic Origin-destination Matrices

3 Modelling in transportation planning

3.1 Choice behaviour

3.2 Model concepts and structure

3.2.1 Model structure: tactical and strategic models

3.2.2 The individual modelling approach

3.2.3 Disaggregation of travel demand

TRAIL Research School, Delft, December 1999

3.3 O/D matrices and transportation modelling

3.3.1 Dynamic O/D matrix estimation

3.3.2 O/D matrix estimation: an overview of techniques

3.3.3 Structure of O/D matrix estimation

3.4 Summary and conclusions

Author
Drs. Ch.D.R. Lindveld
Faculty of Civil Engineering and Geosciences, Delft University of Technology

Thesis supervisor
Prof. dr. ir. P.H.L. Bovy
Faculty of Civil Engineering and Geosciences, Delft University of Technology


## Contents

### Abstract

1  Modelling and estimating dynamic origin-destination matrices .......................... 1

1.1 Introduction ....................................................................................................................... 1

1.2 The case for further work in O/D matrix estimation ......................................................... 1

1.2.1 Relevance of transportation modelling ...................................................................... 1

1.2.2 Expected benefits from further work in O/D matrix estimation ................................. 2

2  Using models to describe object systems ......................................................................... 7

2.1 Subject matter, models, and solution techniques ........................................................... 7

2.1.1 Interactions between the levels .................................................................................. 12

3  Modelling in transportation planning ............................................................................. 14

3.1 Choice hierarchy .............................................................................................................. 14

3.2 Modelling: objects and objectives .................................................................................. 16

3.2.1 Model structure: tactical and strategic models .......................................................... 17

3.2.2 The traditional modelling approach ........................................................................ 18

3.2.3 Disaggregate demand modelling ............................................................................. 19

3.2.4 The market equilibrium modelling approach ............................................................ 20

3.2.5 The dynamic system modelling approach ................................................................ 21

3.3 O/D matrices and transportation modelling .................................................................... 22

3.3.1 Dynamic O/D matrix estimation ............................................................................. 22

3.3.2 O/D matrix estimation as demand model calibration ............................................... 23

3.3.3 Structure of O/D matrix estimation ....................................................................... 23

3.4 Summary and conclusions ............................................................................................. 27

4  References ........................................................................................................................ 28
Abstract

The subject of this paper is the role of Origin-Destination matrix estimation seen from the perspective of demand model calibration. The aim of this paper is to gain a clearer understanding of the mechanisms that determine the demand for transportation modelling by considering its nature and potential applications.

This led towards the idea to investigate the role of O/D matrix estimation in transportation modelling, and to examine the relevance of further work in O/D matrix estimation. To do this, a look was taken at model building per se, and transportation modelling in general.

Although, admittedly, the importance of transportation planning is limited by the chaotic aspects of the transportation system, and the cost/benefit ratio of increasing the level of detail of current transportation modelling practices is a subject for discussion, there seems to be a good case for dynamic O/D matrix estimation.

It will be argued that O/D matrix estimation has close links with the calibration of transportation demand models, and that in some respects O/D matrix estimation can be interpreted as estimating parameters in demand models. Although this approach entails the risk of forcing possible ill-fitting models to observed data, it vastly reduces the amount of data needed to estimate and use dynamic O/D matrices.
1 Modelling and estimating dynamic origin-destination matrices

1.1 Introduction

The subject of this paper is the role of Origin-Destination matrix estimation as seen from the perspective of demand model calibration. The aim of this paper is to investigate the role of O/D matrix estimation in transportation modelling, and to examine the relevance of further work in O/D matrix estimation. To do this, a look was taken at model building per se, and transportation modelling in general.

The paper starts with a short discussion of whether doing additional work on O/D matrix estimation can be justified, followed by a short note on model building in general. In section three, some approaches to transportation modelling are presented. In section four, static O/D matrix estimation methods are reviewed, placed in the context of transportation demand modelling. Next a number of approaches to dynamic O/D matrix estimation are presented. Finally a framework is proposed for existing O/D matrix estimation, which should be applicable to a variety of situations.

1.2 The case for further work in O/D matrix estimation

The estimation of Origin/Destination matrices in transportation modelling is hardly new; people have been doing it without fundamental changes for about 20 years now, so how can we justify further work in this area? Furthermore one can have serious doubts as to whether the quality of the O/D matrix (or indeed of any of the transportation models used) makes any difference when it comes to formulating transportation policy in general, and more specifically when major investment decisions are to be taken.

1.2.1 Relevance of transportation modelling

To deal with the second question first: even the best and the most advanced transportation models known to date are simplifications of reality, depend heavily on exogenous variables (often in the form of scenarios), and must needs be applied to external data whose quality and level of detail can be disputed (owing e.g. to constraints on budget, time, processing capacity etc.). Also it is recognised that in reality the future is far less "computable" than transportation models might suggest.

Furthermore any major investment decision affects interest groups, and effectively reflects the outcome of negotiations between interest groups. Whilst transportation modelling may provide a framework for an orderly discourse, the processes of "bartering between interest groups", and "reconciliation of opposing interests and viewpoints into a workable solution" are generally known as "politics", and "government" rather than as "transportation planning".

TRAIL Research School, December 1999
The best contributions that transportation planning as a discipline is able to make seem to lie in providing understanding of and insight into the transportation aspects of society, and establishing truthful and reliable cause-effect relationships, rather than cut and dried solutions. Transportation models may serve as a "prism" through which policy measures may be ranked according to their expected effects. The valuation of the expected effects will likely be carried out elsewhere, by others, and with other tools than those of transportation planning.

1.2.2 Expected benefits from further work in O/D matrix estimation
The first question might best be interpreted as an inquiry into the expected benefits of continued work on O/D matrix estimation. After all transportation model systems used in practice for what-if analyses rely on skilful approximations throughout, so who is to say that work on O/D matrix estimation can lead us forward?

This question has several aspects, first of all there is the question of internal consistency of the models, secondly there is the question of external consistency, and finally there is the question of the required level of detail.

1.2.2.1 External consistency
By external consistency one usually means that observable quantities in the model should match those observed in reality. Origin-Destination matrix estimation is concerned with ensuring external consistency of a key part of any transportation model: how many people are trying to get from where to where?

In the past, when market equilibria were less explicitly modelled in transportation models, the model components were typically run in succession, without consistency checks. When the policy issues to be addressed by the model system can be discriminated well enough by simplified transport models (as is often the case in developing countries) modelling of feedback loops may not be needed, and model consistency may not be an issue. In such cases an O/D matrix may be obtained by simply calibrating a gravity model on traffic counts.

In contemporary developed societies however, transportation systems often operate close to their physical limits and the room available to change those limits by policy measures is more limited, so that a finer level of modelling is needed to discriminate between measures that can be taken. In such a context the correct representation of existing feedback loops and constraints is often essential for forecasting models in order to gain credibility.
1.2.2.2 Internal consistency

By internal consistency one usually means that the models constituting the model system don't contradict each other.

At first sight this seems trivial, until one realises the existence of links between the transportation market, the housing market, the land-use market, the activity market, and the existence of feedback loops within the transportation system itself. If a what-if model is to be considered credible, it must either in some way model the most significant of these linkages and feedback loops, or prove them to be insignificant. Various aspects of the feedback loop correspond to different components of the model system, and only by a suitable calibration of the parameters in these components can it be ensured that models correctly reflect both currently observed "levels" in the transportation system and observed sensitivities (e.g. value of time) as a fixed point in the model's state space.

In other words: any model that is fed with data reflecting the current state, should also return data reflecting the current state. Due to the feedback loops in the model system, this has consequences for both the levels and the first-order derivatives of the model system (closely related to its elasticities).

At this point a technicality enters the game: O/D matrices are not generally determined by the (readily observable) link flows. Indeed, many matrices usually exist that give rise to the same flow pattern. However, such matrices can differ profoundly in less-readily observable characteristics such as structure. Secondly the elasticities that other model components will predict may be sensitive to this model structure. Thirdly many of the possible matrices may violate constraints that are less readily observable (e.g. triplength distribution, total fuel expenditure) In the fourth place, the current state of the art is disaggregate transportation models, where e.g. travel purpose is explicitly modelled.

1.2.2.3 Level of detail

Regarding the level of detail of the matrix one can think of geographic detail, but also the degree to which structure of the matrix is correct, the level of detail in the composition of the matrix, and of late its time profile.

Except for the relevance of various aspects of the structure of the O/D matrix, there would be little reason to prefer one matrix over another (provided it's flows match the counts). However, in a recent article by [Van Vuren et al. (1999)], which investigated the effects of dynamic traffic assignment in the NMS (the Dutch National Model System, a comprehensive passenger transportation model system) context, it was pointed out that the assignment methods used were quite sensitive to the departure-time profile used. This poses an additional challenge in O/D matrix estimation.

TRAIL Research School, December 1999
A further technicality is that it is practically infeasible to directly observe the O/D matrix in any detail. When estimating dynamic O/D matrices, this problem is further aggravated by the need to obtain the correct departure time distribution (while respecting all other structure of the matrix). At this point a possible solution seems to be to consider O/D matrix estimation simply as part of the entire calibration effort of the current generation of transportation models, and use available data to identify parameters that determine both the structure and the level of the O/D matrix.

1.2.3 Monetary value of transportation planning forecasts

As pointed out by an anonymous referee, the above discussion should be balanced by considering the cost of erroneous planning forecasts.

In Operations Research, an usual way of quantifying the value of information (regardless of its source) is to calculate the expected difference in payoff of the decisions taken on basis of the information.

Suppose that we have to decide between two possible courses of action: build or don’t build a certain facility. The cost of building the facility is known, but the payoff of the facility depends on the outcome of future events: 01 (facility is needed) or 02 (facility is not needed), which occur with probabilities p1 and p2. The payoff depends on the action taken and the outcome of future events, and can be summarised in the following matrix:

<table>
<thead>
<tr>
<th>Action taken</th>
<th>Outcomes</th>
<th>Expected payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O1</td>
<td>O2</td>
</tr>
<tr>
<td>Build facility</td>
<td>f11</td>
<td>f21</td>
</tr>
<tr>
<td>Don’t build facility</td>
<td>f12</td>
<td>f22</td>
</tr>
</tbody>
</table>

Table 1: Payoff matrix without information

The expected payoff R of building or not building the facility can be estimated as:

\[ R_1 = (p_1f_{11} + p_2f_{21}) - C_1 \] (we build the facility) and

\[ R_2 = (p_1f_{12} + p_2f_{22}) - C_2 \] (we don’t build the facility), provided of course we can estimate the probability of the outcomes O1 and O2.

Now suppose we can get information on the probability of the outcomes, e.g. from a planning study. The planning study either predicts outcome O1 (denoted as I1), or it predicts O2 (denoted by I2). We know that the information is neither perfect nor totally without value, and we are fortunate enough to be able to guess at the probability of the study giving the correct answer (i.e. O1 happening conditional on information I1: O1|I1 etc.).

How much should we be prepared to pay for such information? The standard answer is to construct a second table incorporating the information from the planning study:
The expected payoff of building or not building the facility \( R_{i|i} \), can be estimated as: 
\[ R_{1|1} = (p_1 | I_1 f_{11} + (p_2 | I_1 f_{21} - C_1 \text{ (the outcome of the study is } I_1, \text{ and we build the facility) \ and } \]
\[ R_{2|2} = (p_1 | I_1 f_{12} + p_2 | I_2 f_{22} - C_2 \text{ (the outcome of the study is } I_2, \text{ and we don't build the facility).} \]

The expected value of the information \( EV \) is the difference between the payoffs of building and not building the facility, weighted by the probability of the study outcome:

\[ EV = \sum_{k=1,2} P(I_k) (R_k | I_k) - R_k \]

### 1.2.4 Summary

Given that the cost of building new infrastructure is usually in the order of tens to thousands of millions of guilders, and that the cost of carrying out even the most comprehensive studies seldom exceeds 2 million guilders, having good transportation planning forecasts seems cost-effective.

On the downside, the crucial ingredients are the difference in payoff and the probability of the study giving the right answer.

**Payoff**

For decisions on government policy, the difference in payoff can be complicated by considerations such as which external costs and benefits to take into account, and in what way. For applications in the private sector (e.g. BOT (Build Operate, Transfer) projects) the situation is less complicated in that the only consideration of interest is revenue. It is conceivable for instance that building a new railroad is quite indefensible from a financial point of view, while still being worthy of consideration when all external benefits are accounted for.

**Probability of being right**

Many transportation studies have input variables that reflect the development of the economy, land-use, population composition, etc. In addition transportation studies typically don’t allow for the emergence of revolutionary new transport technologies, or a particular transport mode (e.g. scooter) suddenly falling into disfavour.
Usually such factors are a) recognised as sources of uncertainty b) exogenous to the transportation planning model, and c) incorporated through scenarios. The upshot is that the risky task of putting probabilities on the scenarios is devolved onto the users of transportation planning studies.

In view of the above considerations transportation planning still has a role to play, which provides justifiable cause for advancing the state of the art (and ultimately the state of practice).
2 Using models to describe object systems

This section contains a brief digression on the use of models to get a grip on systems under study, with examples from their use in transportation planning.

2.1 Subject matter, models, and solution techniques

Modelling takes place in a "domain of discourse", on objects which are identified within this domain, often by collecting phenomena into convenient bundles. Models are then formulated to describe relationships between objects (the model components). In this way models both reflect which objects an observer distinguishes, and which interactions receive attention. In attempting to capture the essence of phenomena in a domain of discourse models simplify reality.

The most obviously successful models are mathematical (especially quantitative) models, because this type of model permits one to a) calculate results from the models, and b) use mathematical techniques to arrive at inferences. This has proved extremely successful since mathematics works as an intellectual crowbar. It also enables one to spin long but reliable chains of inference which can be analysed and discussed in detail. An often-heard criticism is that with this approach one can lose sight of the subject matter, and bury oneself in technicalities.

Nonetheless some understanding of the domain of discourse is needed for the construction of a (mathematical) model so that an understanding of a domain of discourse usually precedes mathematical modelling. In the remainder of this paper it will be useful to distinguish between the domain of discourse, the modelling level, and the operational level of the model with its assorted tools for doing calculations.

An insightful visualisation of the relationship between subject matter, models, mathematical models, and solution techniques for mathematical models was taken from [Van Nes (1999)], and is reproduced in Figure 1.
The visualisation shows three levels:
(I) the Domain of discourse, which contains the objects and phenomena under consideration
(II) the Modelling level, which contains the models
(III) the Operational level, which contains model implementations and tools for performing calculations on them.

Domain of discourse
With respect to identifying objects in the domain of discourse, the current set of concepts in the study of transportation systems seems to be fairly rich. An example of objects in the domain of discourse are trips, tours, and activity patterns. The extent to which objects and concepts can be represented in models varies, since the complexity of modelling (and implementation!) generally seems to increase much faster than the conceptual complexity.

For example, modelling traffic demand in terms of trip-chains instead of trips constitutes only a small increase in the conceptual complexity. Modelling trip-chains however leads to a dramatic increase in modelling complexity because trip-chains may have a primary and a secondary destination. The attractiveness of the primary destination (e.g. workplaces) is influenced by the accessibility of attractive secondary destinations (e.g. shopping centres). The attractiveness of shopping centres is co-determined by the attractiveness of nearby workplaces.
Furthermore, one may have to model the choice of secondary destination in a trip-chain of 3 destinations conditional on the location of the origin and the destination, leading in principle to $N^3$ alternatives that need to be evaluated, which is a serious problem in real-world applications (although this problem can be reduced by taking account of the "action spaces" of travellers).

**The modelling level**
At the modelling level, various types of model can be distinguished, as shown in Figure 2.

![Figure 2: types of models](image)

Let us call every orderly mental map of reality a conceptual model. Conceptual models are so common in human thought that they are difficult to recognise as models.

**Mathematical models**
A special type of conceptual models are mathematical models. They differ from other models in that they are formulated in terms of mathematical relationships between objects. Such models can be non-quantitative (e.g. catastrophe theory applied to the occurrence of queues), or quantitative. One of the biggest advantages of mathematical models is that they separate the abstract structure of a problem from its appearance. This greatly facilitates accumulation of knowledge because results obtained elsewhere which pertain to the same mathematical structure can be applied directly to the problem at hand.
Quantitative models

Quantitative models can be used in numerical calculations; this is crucial when some of the effects under consideration counteract, or simply when the number of interacting objects becomes larger than 3.

Quantitative models often take the form of "fitted curves", regardless of causal relationships. This can cause serious problems. One of such is known as the "ecological fallacy" (see [de Donnea, 1971]), in which undue aggregation across discrete groups can completely change the correlation coefficient. Only a lack of regard for the underlying mechanism of destination choice could have resulted in this aggregation. Another problem is that fitted curves are often fine for interpolation, but not for extrapolation. Examples of quantitative models in transportation analysis are: regression models for trip production, the gravity model for trip distribution, V/F curves (curves that relate the mode split proportions to the ratio of car travel-time and public-transport travel time), the PCU factor, the application of price elasticity coefficients etc..

Complete models

A model can be called "complete" if it covers all important aspects (i.e. the complexity) of the phenomenon being modelled, regardless of the question whether it does so in a correct way.

Causal models

Although no single definition of what causality means seems to be universally used, three requirements for x to be the cause of y are: isolation, association, and direction of influence (see [Bollen (1989)])]. The interpretation is that y should be isolated from all influences except x, a change in x should accompany a change in y, but not the other way round.

One can expect that causal models have a better chance of being complete models than non-causal models. Furthermore, that satisfy these properties, can be applied and adapted to a specific situation with more confidence than other types of model. Also causal models can be expected to be more transferable than other models because of the isolation property, and more amenable to concatenation.

Quantitative causal models

For the purpose of what-if analyses, causal have models obvious advantages: a (correct) causal model contains the most important relationships between the objects being modelled, and can lay claim to a certain measure of completeness: what is not modelled is not important. Therefore it is often easier to judge if and to what extent quantitative causal models can be applied outside the parameter area in which they were calibrated. Since one of the key issues in transportation planning is producing "what-if" analyses, there is a strong preference for quantitative causal models.

Note that one can derive non-causal quantitative models from a quantitative causal model simply by holding one or more of its inputs constant while letting the others free. This results in a quantitative model with a limited area of validity.
This type model is encountered in practice when a large complicated model is used to calculate elasticities for a certain market equilibrium, which are then used as a first-order approximation model. Sometimes such models are packaged and sold as a separate product.

**Identifiable models**

As soon as a quantitative mathematical model has been formulated, the next difficulty is "model identification", i.e. determining its free parameters (if any), and thereby distinguishing it from other models with the same mathematical structure. This leads us to a further subclass of quantitative models: identifiable models.

A problem that has traditionally beset modelers using quantitative causal models is the issue of obtaining statistical estimates of a model's free parameters on basis of available data. Often the available data determines which models can be estimated, so that identifiability of a model has both pragmatic and fundamental aspects. Identifiability can usually be taken to mean "identifiable with data that could be obtained with current techniques".

**The operational level**

Often quantitative models in transportation analysis operate on too much data to make model application tractable by a pen-and-pencil approach. In such cases the model has to be incorporated in a computer program for application. Unfortunately the identification of the model with the computer code can be so complete that it limits conceptual thought about models.

These issues present a final barrier against model improvement, which in practice turns out to be a surprisingly effective one.

As noted in [Ben-Akiva (1992)], the state of practice in transportation planning has increasingly fallen behind the state of the art over the past 20 years, with which the authors mainly refer to the use of aggregate models instead of disaggregate choice models. Partly because of the need for new skills required to build and operate the models, and partly because of the need to gain credibility with decision-makers.

Part of the reason for the lag may also be a lack of understanding of the advantages of more sophisticated models, and sometimes by of sheer inertia. After all, it is easier and convenient to use an off-the-shelf modelling package, and to simply restrict the choice of available models to what is supported by the software, than to actively seek out the best modelling method.

Another factor are budgetary constraints: simple aggregate models may fall within budget, whereas more sophisticated models may not. If the strengths of disaggregate models (prediction capabilities) are not needed, aggregate models may be used. Admittedly in application, disaggregate models may need to be calibrated before they reproduce aggregate traffic flows with precision. Pivot-point models can be used as a way of coping with such problems.
2.1.1 Interactions between the levels

The diagram in Figure 1 suggests several relationships and feedback loops between levels, to we will now turn our attention.

Top-down relationships
Level (I) to level (II): what is being modelled is reflected in the model structure, and some knowledge about the domain of discourse is needed before one can begin modelling. Furthermore the accuracy with which the study objects can be observed puts natural limits on what should be modelled.
Level (II) to level (III): the desire to operationalise existing models is a strong driving force in the development of mathematical techniques and computational facilities. Examples of the latter abound: e.g. CAD packages, fluid dynamics, financial modelling, meteorology, statistics.

Bottom-up relationships
Level (II) to level (I): which mathematical models can be built strongly depends on what mathematical tools are available. Also, available techniques tend to shape the paradigm people apply in the domain of discourse (the application of mathematics and statistics to engineering, social sciences, and earth sciences is a good example). It can be called the "I've got a hammer" effect.
Level (III) to Level (II): the ease with which what models can be made operational (or the lack thereof) determines which ones can be put to practice (and calibrated). This strongly determines which models attract people, funding, and hence further development.

Positive feedback relationships
Interaction between level (I) and level (II): normally the model-building process forces researchers to revise their views and concepts of the domain of discourse (extreme examples can be found in physics), resulting in changes in the models being built.

It turns out that understanding of (quantitative) mathematical models can be greatly improved by numerical experiments. Traffic assignment to study the effects of infrastructure changes is a case in point. Further examples abound: simulation (O.R. and logistics problems), statistics (resampling, markov-chain monte-carlo), partial differential equations (fluid flow problems), optimisation problems.

Even more importantly, consequences, properties mis-specifications and limitations of model formulations are often recognised only after attention is drawn to them by numerical experiments. Examples are: chaotic behaviour of dynamic systems, the relationship between dynamic and static traffic assignment as reported in [Van Vuren et al. (1999)].
In this respect we would like to point to the importance of visualising model results obtain from numerical experiments (e.g. plots, GIS). We can note that numerical experiments tend to form an integral part of the development process of quantitative models. It is this mechanism which reflects the practical experience gained by practitioners in the field of transportation analysis.

Figure 3: bottom-up relationships

Negative feedback relationships
The converse of the positive feedback effects also exist, and are perhaps less often realised. Serious difficulties in implementation or model-identification problems can condemn otherwise valid modelling approaches to obscurity.

For example, the notion that activities underlie travel demand is obvious, and was first voiced in the sixties. However the field did not receive much attention before it was realised that conventional models fail to account for certain relationships, such as trip chaining, the effect of changing opening times, and general changes in the structure of social interaction (such as more flexible working hours, increased labour participation of women, increased individualism, less cohesion within families etc.). At the moment, quantitative activity-based transportation modelling is largely confined to the academic environment.
3 Modelling in transportation planning

In transportation systems we see various interacting and counter-acting forces and feedback loops. To estimate the net outcome of this interplay of forces on the scale of a city, region, or country, one has to resort to quantitative modelling. Furthermore, due to the spatial heterogeneity one usually needs to take account of a mass of input data that makes computation by-hand infeasible.

A basic aspect of transportation models at this scale level is that they model the transportation system as a reflection of the individual trips of many individuals who can exercise a large mount of free choice, but who affect each other through their actions. Furthermore the individuals, are modelled as having full information about the system they are in, and behaving in rational ways to fulfill their transportation needs (i.e. that they choose a strategy that minimises their time, cost, and effort w.r.t. the transportation system).

3.1 Choice hierarchy

In transportation systems, certain choices are changed more frequently than others (e.g. route choice, and departure time are changed more frequently than mode choice and destination choice) and some need a stronger stimulus to change (e.g. mode choice is less changeable than route choice). In such a system it looks as if several choice strategies are being used and optimised simultaneously, but on different time scales. This has been noted in transportation analysis, leading to the notion of a choice hierarchy.

Usually the following structure of hierarchical choices is adopted to model decisions resulting in travel demand, in which a distinction is made between long-term decisions, mid-term decisions, and short-term decisions. Individual decisions are then modelled through choice models. An example of such a hierarchy is presented below (see: [Ben-Akiva en Lerman (1985)]):

**Urban development decisions (long-term):** Locations of jobs, location of housing

**Mobility decisions (mid-term):** Number of workers, workplace, residential location, housing type, car ownership, mode to work

**Travel related decisions (short-term):** Frequency, destination, mode, route, departure time

The point is that although all decisions interact, due to the difference in updating frequency of the short, medium, and long-term decisions, the short-term decisions can be modelled on their own with the medium to long-term decisions acting as boundary conditions. This is reflected in Figure 4, with the long-term decisions at the top and the short-term decisions near the bottom. The result of these collective decisions is reflected in the state of the transport system and hence it’s level of service, which in turn has an impact on each of the decision levels.
The separation between the choice levels is born out by a study into the opening of the Amsterdam Orbital Motorway, which reports considerable changes in departure time choice and route choice, small changes in mode choice, and practically no change in destination choice and trip frequency (see [Ministry of Transport and Public works (1992)]).

Although the decision hierarchy is generally accepted, it is interesting to note that it implies that the current day-to-day state of the transport system is determined by decisions with widely varying time-scales. This has a number of consequences.

The first is that given the continuous changes in e.g. land use, housing prices, emergence and decline of neighbourhoods, network level of service, working hours, and the life cycle of people themselves, one might wonder if the transport system can at any time be considered to be in equilibrium. The obvious answer is that feedback loops based on short-term decisions can reach equilibrium far quicker than those related to long-term decisions, so that travel-related activities may be in equilibrium while those related to life-style and/or urban development may not be.

Secondly, if one were to regard the system as a collection of coupled feedback loops, and if one were also to suppose that the system is not in equilibrium, then one would expect to see a mix of frequency responses from the feedback loops.
This is one of the reasons for distinguishing between strategic (long term), tactical, and operational (short-term) models. For any modelling level, the processes with a longer time horizon would be incorporated as fixed constraints, while the effects of the shorter-term processes would be regarded as "noise", to be removed by aggregation and filtering.

### 3.2 Modelling: objects and objectives

As noted earlier on, quantitative models in transportation analysis are used primarily for what-if analyses. The fact that the object system itself has dynamics in several different time scales has implications for what time-scales need to be considered when addressing a what-if question, what objects in the domain of discourse are relevant, and in what level of detail they should be examined. This has consequences for any models used to support a what-if analysis.

*Time-scales for dynamics*

The passage of traffic through a transport network is an intrinsically time-dependent phenomenon. On a 24-hr basis, the capacity of most road transport networks are ample to accommodate the demand; but capacity shortages occur during the peak periods.

Evidently the performance of the transportation system (and the resulting L.O.S.) is significantly affected by the *within-day dynamics* of transport system. Relevant time-scales vary on the level of physical detail one regards; they can vary from seconds (at the level of individual vehicles) to quarter-hour periods (when considering aggregate traffic characteristics).

On the other hand, traffic volumes on specific spots in the network may change because people change their routes, modes, departure time etc. because of their experience with the transport system in previous trips. These aspects are called the *day-to-day dynamics* of the transport system.

Of course travel demand changes by day of the week, and there are seasonal influences (e.g. holidays excursions, bad weather conditions during winter).

At this point, the application areas of the models can be subdivided into strategic (time horizon > 10 years), tactical (time horizon weeks to 2 years), and operational (time horizon minutes to hours), according to the time-scale and the level of detail that corresponds to the time scale.

*Static approximation*

Traditionally, transportation planning has focused on the "average working day", because it represents the situation that occurs most often, for which it made use of so-called "static" traffic assignment, in which dynamic aspects of traffic are largely ignored.

The state of practice is that one divides the day into several periods (e.g. morning peak, evening peak, and rest-day), apportions the proper average demand to that period, and uses within-period static assignment.
That the static approximation can be made to more or less accurately reflect
dynamic aspects of traffic flow is illustrated in [Van Vuren (1999)].

In this study, Qblock (a static assignment technique modified to account for
back-blocking) was compared to two dynamic traffic assignment methods
(using 30-minute time slices). It proved possible to adjust the dynamic
assignment methods to be consistent with the static one; under these
circumstances aggregate quantities such as total travel distance, total travel
time, percentage of time lost, and average speed were similar for all three
models. A quick comparison with observed congestion spots however showed
large discrepancies between all three models and reality. The study noted
however that the dynamic models may not have been able to show their
strength because of the 30-minute time slices, and that the quality of the O/D
matrices strongly influenced the results of the dynamic assignment methods.
This provides some motivation for the study of dynamic O/D matrix estimation.

*Common properties*
What all these analyses have in common is that they rest on a quantitative
assessment of 1) "how many travellers go" 2) "from where to where", 3) "by
which mode", and 4) "using which route". In other words, these four quantities
are necessary (but not always sufficient) statistics for the objects in the
domain of discourse (the transport system) that most of the what-if questions
that transportation modelling is currently used for relate to.

As a result, these modelling efforts concentrate on providing estimates for
these quantities.

**3.2.1 Model structure: tactical and strategic models**
For tactical models, we are likely to face the need to incorporate within-day
dynamics. For strategic models the need is less urgent, but may be present if
the transport demand is sharply peaked, since static assignment is a typical
example of a fitted-curve non-causal quantitative model. As reported in [Van
Vuren (1999)], dynamic modelling takes a lot more work, and is not
guaranteed to be correct either, but at least it explicitly captures some of the
mechanisms that are present in the passage of traffic through car networks.

For such models, the familiar structure shown Figure 5 has evolved to capture
the decisions listed in Figure 4 that determine the four quantities to be
estimated ("how many people from where to where by what mode and what
route").
Shown in the top row are inputs to the demand model such as land use (where do people live, work, and play), population characteristics (how many people are there, and what sort of people), and level of service (how much time and effort is it to travel). The travel demand output by the demand model is typically represented as a (synthetic) O/D matrix. A separate traffic assignment module estimates which routes will be taken, and what traffic flows will result on the transportation network.

The demand model corresponds to aspects “how many”, “from where to where”, and “by which mode”, and the traffic assignment module models the “by which route” aspect. Since the cost (time, money, discomfort), as represented by the “level of service” of transportation system has an impact on transportation demand, a feedback loop is usually modelled, as shown in Figure 5.

A number of feedback loops can be accounted for in the single feedback loop shown above, depending on which parts of the choice hierarchy are modelled.

### 3.2.2 The traditional modelling approach

The traditional transportation modelling approach (which is still the state of practice!) is the four-step method (trip generation, distribution, mode choice, traffic assignment).

This approach has been criticised for two main shortcomings:
- lack of consistency between modules, because it does not model feedback
- prediction errors at any stage are compounded in subsequent stages.

A common response is to attempt to achieve consistency between the above modules by feeding the results from traffic assignment back into the trip generation stage. Although this approach may work in practice, it still does not guarantee convergence to a consistent solution.
Aggregate demand modelling

The first three steps of the four-step model deal with demand modelling. In traditional demand modelling, aggregate demand modelling (which assumes all persons to behave in the same way) is the oldest, most widely known, best supported, and perhaps most often used modelling approach.

Severe criticism has been levelled at it for the past 20 years, and it is still with us. Examples of criticism are resistance to mode-shift (captives), VOT issues. An example of problems that may arise in aggregate models is known as the "Ecological fallacy"; see [de Donnea (1971)].

3.2.3 Disaggregate demand modelling

In response to the criticism directed at aggregate demand modelling, researchers have developed disaggregate demand modelling, which takes account of the heterogeneity of users of the transport system, such as licence holding, car ownership, age, income class, household composition, travel purpose etc.. Classes of users are distinguished which are as homogeneous as possible with respect to the travel decisions they make. The population is divided into such classes, and in model application the decisions situation for each class is treated separately.

Such models require more skills to estimate, implement, and use than aggregate models, but they allow planners to avoid some serious pitfalls.

Discrete-choice modelling

One of the mainstays of today's transportation models are discrete-choice models, and especially models based on random utility theory (see [Ben-Akiva et al. (1985)]).

In discrete-choice models based on random utility theory choices are characterised by their attributes, and have a utility, which is measured on an interval scale. This type of choice models are compensatory choice models in the sense that they focus on trade-offs between attributes. Other models exist (see e.g. [Garling, Laitle, and Westin in Garling et al. (1998)]) such as bounded rationality models.

The activity-based approach

One of the criticisms levelled at discrete choice modelling based on utility theory is that it assumes that individuals are utility maximisers. This is a strong assumption, which may not always be justified. Alternative approaches include that of bounded rationality (see [Hato et al. (1999)]).

A relatively new approach emphasises modelling peoples' activity patterns, and deriving travel demand from that. This is especially important when studying secondary trips travel induced by discretionary activities (such as shopping, sports, social visits etc.), since in such cases one lacks the structure imposed by mandatory activities such as work and school. Even then, the discrete choice modelling framework may apply, albeit to other types of choice (e.g. activity patterns).
3.2.4 The market equilibrium modelling approach

As soon as individual models in the 4-step approach are joined together, the idea of requiring consistency between the model components is an obvious one. This consistency can be realised by viewing the transport system as a series of coupled "markets", and requiring a market equilibrium each of these markets separately, and for the lot of them together.

Although the term market equilibrium has intuitive appeal, it is a concept that primarily corresponds to idealised markets, where supply and demand curves are continuous monotone functions of costs. As shown in economics, this is a strong simplification of the state of affairs in real-world markets (market heterogeneity, transparency, full competition etc.) but a useful one.

![Market equilibrium in an idealised market](image)

Although market equilibrium can be applied to large transportation systems, the underlying hypotheses have been criticised. The most often criticised hypotheses are:

- Market transparency, which is unreasonable considering that the users are individuals with limited information acquisition and processing capabilities.
- Economic motives as the only explanatory mechanism; trade-off possible between all aspects of alternatives
- Existence of a stable market equilibrium, which ignores time-dynamic aspects. This assumption is problematic because both individual and environmental conditions change day by day, and market may be far from transparent.

Range of applications

Market equilibrium approaches have a very wide range of application: the majority of all choices that travellers make can probably be interpreted as behaviour in a certain market situation. Discrete-choice models in particular fit in very well with the notion of market equilibrium, since they are typically sensitive to variables that are determined by the outcome of the collective choice.
Examples are static and dynamic traffic assignment (see e.g. [Chen (1998)]), mode choice, departure time choice, residential location choice etc.. The often criticised assumptions of total market transparency can be mitigated e.g. by allowing for perception error on part of the users ("Stochastic User Equilibrium"), and by allowing for information processing capability, and bounded rationality (see [Hato et al. (1999)]).

Tools
Variational Inequality (VI) models have emerged as a suitable framework for the formulation, theoretical investigation, and calculation of user-optima (and system optima) of the multi-player market equilibria implicit in network models.

Traffic assignment was cast into VI form about 10 years ago. Other examples are spatial equilibrium models (land pricing), and the choice hierarchy (frequency, destination, mode, departure time).

A hypernetwork approach permits many choices to be modelled as network flow problems, for which theory and solution algorithms already exist. For example, in [Chen (1999)] an example is given of a combined dynamic route-choice, mode-choice and departure-time choice model. A formulation for combined travel, mode, destination, and route choice based on a utility maximisation framework is given in [Oppenheim (1995)]. This gives existence and uniqueness theorems for the combined models for route choice etc.

3.2.5 The dynamic system modelling approach
As noted before, a number of feedback loops can be distinguished in transportation systems, and the route-costs in a transportation system in response to travel demand show considerable variance with increasing traffic loads. Under the circumstances, and in view of moderate success rates in predicting the behaviour of transportation systems, the question has been raised if and to what extent transportation systems are chaotic. The term chaotic is interpreted here as meaning that the system's evolution is so sensitive to initial conditions that even the presence of ordinary background noise makes it impossible to predict the system's state within the time horizon of interest.

This situation has led researchers to study transportation systems as dynamic models. This approach is taken e.g. taken by [Cantarella, G., Cascetta, E. (1995)] in the study of day-to-day dynamics of the transportation system, and by [Nagurney, A., Zhang, D. (1998)].

Day-to-day dynamics may be introduced even if we assume the existence of within-day market equilibria. This leads to a dynamic sequence of markets (see [Cantarella, Cascetta (1995)]). Cantarella and Cascetta show that for the nature of the feedback mechanism they propose, only systems that react strongly to losses and show little inertia show chaotic behaviour.
Despite the belief that dynamic models capture certain elements of transport systems (that static models do not), the adoption of dynamic transport systems has been very limited.

**Chaotic systems**

Even if the transportation system should have a chaotic character, what does this mean? If a system is chaotic this is often taken to mean that this system is "a hopeless prospect for prediction". However, inspection of traffic counts on motorways quickly shows that although the traffic state (speed, congestion level, flow level) at 1-minute level is quite noisy, hourly and daily average traffic counts are much less so. In fact, the time scales of the transport system seem to be different enough so that the longer-term dynamics may be considered sufficiently constant when considering the short-term dynamics.

This opens the door to adaptive modelling using a rolling horizon approach, as was shown in the DYNA project (see [Whittaker, (1991)] and [Ben-Akiva et al. (1995)]).

For longer-term predictions, the chaotic nature of the transport system seems to be less significant than the influence of variables exogenous to the transport system (such as e.g. increasing labour participation of women, family size change, etc. as listed [Gunn and Van der Hoorn (1998)]).

### 3.3 O/D matrices and transportation modelling

O/D matrices and travel costs sit on the interface between demand modelling and assignment. An O/D matrix is summary of the demand, and travel costs are the summary of how this demand affects the network.

#### 3.3.1 Dynamic O/D matrix estimation

In O/D matrix estimation, the O/D matrix is never completely accessible to direct observation (if it were, there should be no need of estimating it), so that an O/D matrix can be compared with observations only indirectly through observable quantities that can be derived from the O/D matrix.

The most well-known and direct of these quantities are network traffic flows obtained through traffic assignment. In traffic assignment the O/D matrix is loaded onto the network, and the resulting traffic flows are calculated. Calculated traffic flows can be compared to observed traffic flows, and the O/D matrix can be adjusted so that the assigned flows best match the observed. In traffic assignment one has to model route-choice through the network, which can introduce extra errors. This problem can be reduced by using counts on screenlines or cordons, where route choice is eliminated.

Other quantities that can be observed more or less directly are the trip length distribution (household survey), row and column totals (household survey), and partial matrices (roadside interviews).
3.3.2 O/D matrix estimation as demand model calibration

In view of the difficulties posed by the fundamental scarcity of observed O/D matrix cells, and the possibility of conflicting datasets one can try to estimate O/D matrices by using part of the available data to calibrate the parameters of the demand model.

In [Cascetta and Russo (1997)], a framework is sketched in which transportation demand in the form of an O/D matrix by travel mode is modelled as a function of a parameter vector $\beta$. The model incorporates generation, distribution, and mode choice in a single choice model with parameter vector $\beta$. The O/D matrix is denoted as $t$, which depends on $\beta$: $t(\beta)$. This parameter can be linked to several observable quantities, e.g. link flows which permit its estimation.

Even though the relationship between observable and unobservable quantities may be somewhat tenuous, the data-density per parameter is much greater than in the case where each matrix cell is treated as an unknown. In fact this type of model may be the only feasible way to estimate large dynamic O/D matrices.

3.3.3 Structure of O/D matrix estimation

A large number of approaches towards O/D matrix estimation and modelling have been documented. For use with strategic and tactical transport models, there are advantages in adopting the structure shown in Figure 7 for both static and dynamic O/D matrix estimation, because the position of the market equilibria can be accounted for in the course of the estimation.

Shown on top is the same transportation model structure as shown in Figure 5. Both the synthetic O/D matrix produced by the demand model and the network flows that result from the assignment procedure can be compared to observable data, and indicate in which respects the synthetic O/D matrix does not match observations.

The mismatch between modelled and observed quantities related to the O/D matrix is quantified in an objective function (third from bottom). By minimising this objective function, and optimal (in the sense determined by the objective function) O/D matrix estimate can be obtained.

The adjustment factors are fed back into the demand model, and in some cases directly modify the synthetic O/D matrix. The synthetic O/D matrix that corresponds to the minimum of the objective function is the estimated O/D matrix.
Note that the objective function is related for the most part to indirect measures of the O/D matrix because the true O/D matrix is not directly observable. Also note that comparison to observed flows is possible only after traffic assignment. This means that O/D matrix estimation is dependent on the correctness of the assignment method; in fact the O/D matrix estimate may change when a different assignment technique is used.

3.3.4 Example of static model-based O/D matrix estimation
As an example of static model-based O/D matrix estimation we can point to the base-matrix estimation of the Dutch NMS (National Model System), as described in [Gunn, Meijer and Lindveld (1997)].
In this estimation exercise, the synthetic matrix was taken from the Dutch NMS, and the following multiplicative model was assumed:

$$\hat{T}_0 = T_0^o \alpha_i \beta_j \prod_{l=1}^{L} \gamma_{lj}$$

with:
- $T_0^o$: the apriori matrix from the NMS
- $\alpha_i$, $\beta_j$: parameters corresponding to production and attraction of each origin and destination zone
- $\gamma_{lj}$: a parameter corresponding to each of the available other datasets (observed partial matrices, trip length distribution, traffic counts)
- $p_{lj}$: the fraction of traffic in the (i,j) cell of the matrix that is captured in dataset l

In the base matrix estimation the observations were assumed to follow a Poisson distribution, and parameters $\alpha_i$, $\beta_j$, and $p_{lj}$ were estimated using Maximum Likelihood estimation. Where the Poisson distribution did not seem appropriate, a variance destabilisation transformation was applied to arrive at a variable with equal mean and variance, which was then treated as if it was Poisson distributed.

This method has several advantages:

- In the estimation procedure about 2000 parameters had to be estimated to fit a matrix of 160000 cells, thereby reducing the number of free parameters by a factor of 80 over methods in which the matrix is estimated directly.
- The structure of the model to makes that the estimation procedure applies the minimum necessary correction to the apriori matrix, thus ensuring that the structure specified by the NMS demand model is overruled only where necessary; this is important because the number of matrix cells is usually so large compared to the available observations that many matrix cells are completely unobserved
- As long as an adequate demand model is available, this type of procedure is particularly flexible in its data requirements since the method focuses its change on cells for which it has observations and affects other cells much more lightly
- The method is well able to cope with and arbitrate between contradictory data. This is particularly important because real-world data practically always contains considerable amounts of noise and observation errors.
3.4 Extension to the estimation of time-dynamic O/D matrices

If a transportation demand model is available for the static case, then a departure time model such as proposed by [Bhat (1998)] can be added to the modelling framework to produce a dynamic O/D matrix, leading to a model of the type:

\[ \hat{T}_{ij} = T_i^o \alpha_i \beta_j \prod_i \gamma_i^{ij} \tau_{ij}, \]

where:
- \( \hat{T}_{ij} \) : the time-dynamic O/D matrix to be estimated with \( t \) a time interval
- \( \alpha_i, \beta_j \) : parameters corresponding to production and attraction of each origin and destination zone
- \( \tau_{ij} \) : the output from the departure time choice model, giving the probability of a trip-make in the O/D cell \((i,j)\) chooses to depart during time interval \( t \).

3.4.1 Related research areas

Currently developments are under way (e.g. in the PLATOS project, which is a joint research effort by the AVV, several universities, and transportation consultants) to develop an operational estimator for EE (Entry-Exit) matrices for motorway corridors.

This work is primarily motivated by the desire to improve the DTC (Dynamic Traffic Control) aspect of DTM (Dynamic Traffic Management) capabilities. The interest in dynamic EE matrices is that it facilitates feed-forward rather than feed-back traffic control. Feed-forward traffic control can avoid certain types of instability.

Such EE matrices can be used to as additional objectives that the dynamic O/D matrix to be estimated should satisfy.

There are several advantages in following this approach instead of directly using dynamic traffic counts:
- the EE matrix estimator can be used to deal with the inevitable noise that is present in the traffic counts
- it permits one to separate the highly variable level of the EE matrices from their much more stable structure
- it reduces the complexity of the data-processing in the estimation of dynamic O/D matrices
- by binding the lower-level EE matrix estimation to the O/D matrix estimation it is possible to provide "guidance" to the EE matrix estimator by supplying it with an apriori EE matrix derived from the apriori static O/D matrix and traffic assignment.
3.5 Summary and conclusions

In view of the feedback effects present in transportation systems, there is a need to model these. The within-day dynamic aspects also seem to merit modelling. In view of the strong influence of the quality of the O/D matrix on the assignment results, there is a need for dynamic O/D matrix estimation.

The market-equilibrium approach seems to be a valid approximation to the level of detail required to assess the average situation (as e.g. required for strategic models), even in the presence of stochastic fluctuations and day-to-day dynamics.

Even though the transportation system may exhibit chaotic behaviour, this does not mean that modelling or prediction is impossible, but it does impose limits to the scope and practical importance of transportation planning.

O/D matrix estimation is very much a part of traditional transport modelling, and is less innovative than exploring the mechanisms underlying travel behaviour such as e.g. activity-based transportation modelling. Although dynamic O/D matrix estimation addresses issues in a field which has progressed only slowly in the past 20 years, it is needed to correctly model within-day dynamics in the framework of tactical or strategic models. However, the cost/benefit ratio of incorporating within-day dynamics into such models does not seem to be entirely beyond question.

Although a significant part of the cost of determining dynamic O/D matrices is in collecting and processing the data required to determine the dynamic O/D matrix, the lack of an easy to use, flexible dynamic O/D matrix estimator must be regarded as a serious obstacle against adaptation, or even full-scale field trials.

Furthermore, the extension of existing demand models with a departure time model and calibrating such models against available observations would demand far less data and represent a much more cost-efficient way of estimating dynamic O/D matrices than an ab-initio approach. Also the data demand would be flexible: one would be able to start with a modest data requirement, and improve the matrix as and when needed on basis of additional data collection. The practical consequence is that it would be much easier to start with an existing demand model system (network, matrix, demand model, traffic counters), and add a departure-time model to it than to build a model from the ground up. A suitable test-bed for the estimation and application of a departure time model is being sought. Ideally a network and a calibrated demand model should already be available.

Last but not least, the availability of an easy to use dynamic matrix estimator with flexible data requirements could significantly improve the cost/benefit ratio and the feasibility of doing dynamic O/D matrix estimation.
4 References


• Van Nes (1999) Personal communication, Delft University of Technology.

Contents

Abstract

1. Introduction
   1.1 Urban Traffic Control
   1.2 Intelligent Agents

2. Agent Based Urban Traffic Control
   2.1 Autonomous UTC
   2.2 Intelligent Agent Based Urban Signal Control System

3. Architecture of an Agent Based Urban Intersection Control System
   3.1 Control strategy model
   3.3 System wide optimization of the control strategy

4. Conclusion

TRAIL Research School, Delft, December 1999

Author
Ir. D.A. Roozemond
Faculty of Civil Engineering and Geosciences, Delft University of Technology

Thesis supervisors
Prof.dr. H.J. van Zuylen and Prof.dr.ir. P. van der Veer
Faculty of Civil Engineering and Geosciences, Delft University of Technology
Architecture of an Agent Based Library Information

Control System

TRIST Research School Data December 1989

Author
N. D. Y. Ricksen

Title of Thesis Submitted in Partial Fulfillment of the Requirements for

Doctor of Philosophy

Preface: Till Now Some Preliminary Results of This Work

Lectures in Cybernetic and Computer Sciences 1969
## Contents

The usability and effectiveness of traffic management and traffic control systems greatly depends on its ability of adapting upon traffic patterns and perturbations. In this research the applicability of autonomous intelligent agents in Urban Traffic Control (UTC) and why these Artificial Intelligent strategies are useful in UTC. We propose a traffic control system that can adapt itself at changing environments. A model, and the simulation and validation of the model is presented in this paper. Of an autonomous urban traffic control system that is needed to gain insight into the classical concept of traffic control. The UTC model in presented the work is primarily in predicting traffic pattern and can extend and respond to traffic conditions in real-time. The UTC model is presented the work is primarily in predicting traffic pattern and can extend and respond to traffic conditions in real-time.

### Abstract

1. Introduction ................................................................................................. 1  
   1.1 Urban Traffic Control ............................................................................ 1  
   1.2 Intelligent Agents ................................................................................ 2  

2. Agent Based Urban Traffic Control............................................................ 3  
   2.1 Autonomous UTC ............................................................................... 3  
   2.2 Intelligent Agent Based Urban Signal Control System ...................... 4  

3. Internal architecture of the ITSA model.................................................... 6  
   3.1 Prediction model .................................................................................. 7  
   3.2 Control strategy model ........................................................................ 8  
   3.3 System wide optimisation of the control strategy .............................. 9  

4 Conclusions ............................................................................................... 10  

References ..................................................................................................... 11
Abstract

The usability and effectiveness of traffic management and traffic control systems greatly depends on its ability of reacting upon traffic patterns and permutations. In this research we investigate the applicability of autonomous intelligent agents in Urban Traffic Control (UTC), and why these Artificial Intelligent strategies are useful in UTC. We propose a traffic control system that autonomously can adapt itself at changing environments. A model, and after verification and validation of the model a prototype / simulation program, of an autonomous urban traffic control system will be used to get insight into the technical -and functional- applicability of autonomous, self adjusting and self optimising systems in traffic control. The UTC model is primarily based on several Intelligent intersection Traffic Signalling Agents (ITSA) and some authority agents. In this way we get an UTC system based on agent technology that can adapt and respond to traffic conditions in real-time - maintaining its integrity and stability within the overall transportation system – and in the meantime get a system that is able to make better use of the capacity of intersections.
1. **Introduction**

Designing, implementing, optimising and adjusting urban traffic control systems involves quite some effort and knowledge. Due to several reasons, changing environments do not always lead to changes in the traffic control units. Adjusting a traffic control unit often is a costly and timely affair. Quite often a traffic control unit is not adjusted for short term road works, nor for changing traffic patterns.

1.1 **Urban Traffic Control**

The effectiveness of (urban) traffic control systems greatly depends on its ability to react upon changes in traffic patterns. When this ability to react becomes an integral part of the traffic control unit, the better it can react to changes in traffic conditions. The hypothesis is that it may be useful to make use of self evaluating and self adjusting traffic control systems and that in urban traffic control there is a market for a system that is able to adjust itself if the environment changes. Intelligent signal control systems must have the capability to optimise the traffic flow by adjusting the traffic lights and coordinate operation between each signal in order to maximise the person and vehicular throughput and minimise delay. For intelligent urban traffic control we ideally need a fully pro-active, real-time traffic control system; anticipating what will happen within the next 15 minutes (Roozemond, 1997). For such a system the control plans used, optimised by some sort of performance criteria, are based on actuated traffic conditions and are updated frequently and possibly even during a cycle.

Intelligent urban signal control systems must thus have the capability to handle unforeseeable changes in traffic flow, such as accidents and optimise the traffic flow by adjusting the traffic lights and coordinate operation between each signal. All current traffic signal control methods are based on feedback algorithms using traffic demand data in the past; the time span between data collection, data transmission, calculation of parameter settings varies from years to a minimum of 5 minutes. These methods are not optimal where traffic demand changes rapidly within a time interval of 1 - 10 minutes, such as in the morning- or evening rush hours. A lot of systems (e.g., MOVA, LHOVRA, UTOPIA, SCATS or SCOOT) are already based on traffic-responsive control; often with offset patterns to handle different situations. In the domain of Urban Traffic Control (UTC) more and more research is directed towards operation and optimisation of individual and linked signal-controlled road junctions based on actual traffic (Saito, et al. (1997), Kronberg, et al. (1993), Vincent, R., (1993)). The basic premise is that existing signal-plan generation tools make rational decisions about signal plans under varying conditions; but these tools have no higher order meta-rules nor traffic prediction models incorporated into the system so intelligent updating of the signal control settings is not possible.
1.2 Intelligent Agents

Agent technology is a subset of distributed Artificial Intelligence (AI). The agent paradigm in AI is based upon the notion of reactive, autonomous, internally-motivated entities that inhabit dynamic, not necessarily fully predictable environments. An agent is "autonomous" to the degree that it decides for itself how to convert data into actions in its efforts to achieve its goals. Autonomy is the ability to function as an independent unit or element over an extended period of time, performing a variety of actions necessary to achieve pre-designated objectives while responding to stimuli produced by integrally contained sensors (Zeigler, 1990). The problem solving component of an intelligent agent can be a simple rule-base system but can also be a neural network, an expert system or some simple (fuzzy) rules.

Multi Agent Systems can be characterised by the interaction of many agents trying to solve a variety of problems in a co-operative fashion. While considerable effort is being devoted to understanding the detailed interactions among a few agents and designing operational Multi Agent Systems that can deal with simple problems, relatively little is known about the global behaviour of these systems when they are scaled up to deal with more realistic problems. One of the difficulties in both the design and the understanding of distributed artificial intelligence systems comes from the lack of central controls, and the ensuing conflicting, uncertain, incomplete and delayed knowledge on the part of the agents. It may be obvious that a provision to find a feasible solution is a necessity for an agent. Besides some Artificial Intelligence (AI), agents should have some additional attributes to solve problems by itself in real-time: understand information; have goals and intentions; draw distinctions between situations; generalise; synthesise new concepts; model the world they operate in and plan and predict consequences of actions and evaluate alternatives. The complicated dynamics of multi agent systems can result in many agents making poor decisions when seen from a global perspective (even though they appear reasonable locally). This may lead to significantly lowered performance and great difficulty in programming the individual agents. Thus, it is important to devise effective decentralised control methods that simplify the global dynamics. There are different possible solutions for this problem: the use of a hierarchy of agents or the use of a blackboard architecture. A common communication protocol is a necessity in both cases, preferably based on a common standard like CORBA. For specific control situations an architecture based on hierarchy seems to be a better solution and closer to current traffic control ideas already implemented in traffic control centres.

The research of the usability of agent technology within traffic control can be split into two parts. First there is a theoretical part where an agent architecture has to be developed and agent technology gets integrated with current traffic control theory. The next, and even more important part, focus on the practical sides like implementation, demonstration and performance.
2. Agent Based Urban Traffic Control

The basic functions of an (automated) Urban Traffic Control (UTC) system are reasonably well known as they are used for quite some time. The particular techniques proposed in this research are experimental and not yet mainstream, especially for such a large, real-time, application. After designing what is necessary to build we need to determine how we are going to build that system. Intelligent Urban Transport Systems may require intelligent agents of many types (e.g. data agents, route finding agents, etc.). Here we only address specific ones dealing with urban intersection control.

2.1 Autonomous UTC

We propose a system that autonomously can adapt to changing environments. In that system we hope to get an Urban Traffic Control system (UTC) that is able to adapt and respond to traffic conditions in real-time, maintain its integrity within the transportation system and make better use of the capacity of intersections. The key aspects of improved control, for which contributions from AI and artificial intelligent agents can be expected, include the capability of dealing with multiple problems and conflicting objectives; the capability of making pro-active decisions on the basis of temporal analysis; the ability of managing, learning, self adjust and responding to non-recurrent and unexpected events (Ambrosino et al., 1994).

The essence of an, demand responsive, pro-active, UTC system can be based on several coupled Agents. An intelligent, agent based, UTC system should be capable of calculating and optimising control strategies, as well have knowledge about the intersection(s). Real-time optimisation only works if there is sufficient quality in traffic predictions, a good choice is made regarding the performance indicators and an effective way is found to handle incidents and one-time occurrences (Rogier, 1999). The agents make decisions on how to control its environment based on its internal data. When necessary an agent can request for additional information or receive other goals or orders from its authority agent. This may prove a necessity to get a better, less local, optimum; e.g. not one single optimised intersection but several adjoining intersections working together, and in this way adapting goals of a higher level. To be acceptable as replacement unit for current traffic control units an agent based system should get just the same or better performance then current systems. An agent based system should be demand responsive (or better : pro active) and should be adaptive during all stages and times. Both slow, but permanent, and long term changes as well as abrupt, incidental occurrences should be handled (Rogier, 1999). Furthermore a good balance between all participants is necessary. For example: getting a more even traffic load on infrastructure may not lead to extra environmental impacts or more traffic load in residential areas. Or due to the extra traffic flow on one link, other traffic flows get more congested.
2.2 Intelligent Agent Based Urban Signal Control System

The essence of an, demand responsive, pro-active, agent based UTC consists of several intersection control ITSA's (Intelligent Traffic Signalling Agent), some authority agents and - not absolute necessary - Road Segment Agents (RSA). Such an UTC system requires: monitoring system of traffic with a data base, a rule- or model-base for evaluation and adjustment, a model of the surrounds and an efficient diagnostic routine for both traffic light operations as well as rule- and parameter adjustments. That UTC should be a demand responsive, pro-active, system based on actual information and adapting to situations. To simplify the system we only look at inner city traffic and even only to controlled intersections with detectors. For simplicity reasons during development and implementation we take only cars into account, but other traffic participant should be integrated into the working model directly after the first model is implemented.

The ITSA makes decisions on how to control its intersection based on its goals, capability, knowledge, perception and data. For a specific ITSA, implemented to serve as an urban traffic control agent, the following aspects are incorporated (Roozemond, 1998):

- The ITSA has goals to accomplish: for instance maximizing traffic flow;
- The ITSA has rules to obey and roles to perform;
- The ITSA decide on what actions to take; helped by its controller, the view and knowledge it has of its environment, its abilities and its state;
- The ITSA has skills and tasks that it can perform depending on situations. The agent solves a problem mostly acting on its own 'feeling' and its knowledge.

Figure 1: Simplified architecture of an, agent based, UTC system.
A more specific example of a simplified architecture of an agent-based, UTC system is given in figure 1. There we have one authority agent controlling several (4) intersection agents, which in their turn manage the intersection controls helped by RSA's. The authority agents' tasks are controlling, coordinating and leading the ITSA's towards a more global optimum. Often an ITSA should sacrifice some performance for the purpose of co-operative behaviour caused by appointment of an authority agent or self-control of ITSA's. An intelligent, agent based, urban traffic control system should be capable of calculating and optimising control strategies, as well have knowledge about the intersection(s). The ITSA can directly influence the control strategy of their intersection(s) and are able to get insight in on-coming traffic via other agents or prediction models. Using all available information the ITSA (re)calculates the next, most optimal, state and operates the traffic lights accordingly after synchronising the control schema with adjoining agents.

An ITSA, capable of controlling or advising in real-time, should perform the following specific traffic control related actions (Roozemond, 1996 & 1998):

- Data collection / distribution (via RSA - information on the current state of traffic; from / to other ITSA's - on other adjoining signalised intersections)
- Analysis (with an accurate model of surrounds and knowing the traffic- and traffic control rules define current trend; detect current traffic problems)
- Calculation (calculate the next cycle mathematically correct)
- Decision making (with other agent deciding what to use for next cycle; handle current traffic problems)
- Control (operate the signals according to cycle plan.)
3. Internal architecture of the ITSA model

For a better overview of the internal ITSA models and functions see figure 2. There the overall architecture is sketched. The actual operation of the traffic lights is left to an ITSA-controller agent. Such an agent is a simple computerised box comparable with current traffic control units; just for operating the traffic lights according to the chosen control strategy. The central part of the ITSA is a control strategy agent. That agent can have several control strategies, such as anti-blocking, public transport priority and other adaptive strategies. There may be communication with other ITSA’s of nearby intersections, RSA’s, the urban traffic control centre, other control centres (bus / route guidance, etc.). The control strategy agent uses the estimates of the prediction model agent which estimates the states in the near future and its own knowledge to evaluate current strategy and build a new control plan. The ITSA-prediction model agent estimates the states in the near future. The prediction model agent gets its data specified and related to intersection and road segments; as an agent that ‘knows’ the forecasting equations, actual traffic conditions and constraints and, in combination with the above given data, future traffic situations can be calculated by way of an inference engine and it’s knowledge- and data-base.

![Figure 2: Actuated control strategy based on agents](image)

Traffic dependent intersection control normally works in a fast loop. The detector data is fed into the control algorithm. Based upon predetermined rules a control strategy is chosen and the lights are operated accordingly. Here we suggest the introduction of an extra, slower, loop where rules and parameters of the prediction model can be changed by the meta model (see figure 2). The prediction model is necessary if we want to achieve pro-active adaptive control. Therefor a traffic prediction model with artificial intelligence is needed. The prediction-meta-model may change the parameter settings of the prediction model. The meta model performs a check if recent predictions were accurate and if not, adjusts, according to its internal rules and knowledge, the parameter settings for the prediction model. A more elaborate explanation of the prediction- and control strategy model and the way we hope to handle the possible local optimisation in stead of global optimisation is given in the following chapters.
3.1 Prediction model

As stated before, the ITSA-prediction model agent estimates the states of the traffic in the near future. The prediction meta model compares the accuracy of the predictions with current traffic and is able to adjust, via its meta model, the prediction parameters if the predictions were insufficient accurate. The prediction model agent is fed by several inputs: vehicle detection system, relevant road conditions, control strategies, important data on this intersection and its traffic condition, communication with ITSA's of nearby intersections and the urban traffic control centre. The prediction model agent gets its data specified and related to intersection and road segments from RSA's. The agent itself has a knowledge base and inference engine. In the knowledge base the different rules with forecasting equations and normal traffic conditions over time are to be found. Also constraints regarding its specific intersection and insight into current (traffic) conditions are known to the agent. With these data future traffic situations can be calculated. The kind of calculus used is effectively a state prediction model: predicting the states on time \( t_j \), given the states on time \( t_j - dt \).

Here we want to elaborate a bit further on the proposed ITSA prediction model. The simplest form of a prediction model is to calculate the average of a number of cycles. A more accurate, but still easy, method is to calculate a weighted and adjusted average of a number of cycles (1).

\[
P_{\text{ITSA,link}} = \frac{\sum_{t=T_{x-1}}^{T} x \cdot \xi_{\text{link}} \cdot N_{C_{\text{link}}}}{x!}
\]

where:
- \( P_{\text{ITSA,link}} \) = expected amount of traffic per link
- \( N_{C_{\text{link}}} \) = number of cars per link
- \( \xi_{\text{link}} \) = correction factor
- \( x = \text{Number of cycles (current = highest)} \)
- \( T_{x} = \text{Time needed for } x \text{ cycles} \)
- \( T = \text{Current time} \)

\[
\xi_{\text{link}} = \xi_{\text{day}} \cdot \xi_{\text{time}} \cdot \xi_{\text{weather}} \cdot \xi_{\text{road cond}}
\]

where:
- \( \xi_{\text{day}} = \text{Correction factor for kind of day} \)
- \( \xi_{\text{time}} = \text{Correction factor for part of day} \)
- \( \xi_{\text{weather}} = \text{Correction factor for weather conditions} \)
- \( \xi_{\text{road cond}} = \text{Correction factor for road conditions of link} \)

The correction factor (2) depends on a lot of secondary data, the most important ones are data on important aspects of traffic conditions: kind of day, part of the day, weather conditions, road conditions, etc. The correction factor is also dependent on the driver / car combination but as these information is often unavailable it is omitted. Normally the correction factors will be near 1, as the current traffic conditions are similar to last couple of cycles, but the correction factor may be altered by the meta model.
Especially the correction factors for weather and time may have a large influence during specific occurrences. To be able to forecast properly in all cases we use both historic data and current data from monitoring devices as input for our dynamic model. Thus, combining the best of both worlds, we are able to forecast travel times in the near future based largely on actuated data as well as travel times for some time ahead (>5 minutes) based on actuated and historic data. In general we can see that the need for actuated data becomes smaller as the needed forecast lies further away. Accuracy and variability of data are important as they are the key element when the outcome of these models will be incorporated in traffic management schema's.

The prediction model gives the estimated number of cars in the next period. The predicted forecast is only valid for a limited time. If extra information is necessary the RSA should be able to give some data. When a RSA is situated between two ITSA's the RSA is able to calculate the time when the cars that leave the intersection at link j, arrive at intersection k. The RSA needs therefore a mean velocity and a correction factor depending on time, weather etc. It is presumed that data on upstream traffic can be used to predict the downstream traffic in a specific time period. Research has shown that models using historic, upstream and current link traffic give the best results (Hobeika & Kim 1994). These models only give forecasts for specific links and don't take other links, possible disturbing that link's traffic volume, into account. Thus a more elaborate model could give better results.

3.2 Control strategy model

The control strategy agent uses the estimates of the prediction model agent to calculate the best control strategy. Therefore some knowledge in the form of production rules are needed: given specific situations the following counter action is appropriate (IF <condition 1><condition 2>...<condition n> then <decision rule 1>...<decision rule m>). Some data has to be fed into the system initially. Data on the intersection, like layout, guaranteed green, yellow and red times and the conflict matrix. Data on average flow on the links can be gained by the system during run-time but may also be initialised. The ITSA control strategy model calculates signal plans to pro-act on the predictions of the prediction model agent. The strategy model checks with other agents if specific actions have to be taken and, in the same process, sends other agents its control strategy. In normal non conflicting cases when only optimisation plays a role, a decision model agents' choice based on a number of criteria with different weights will be sufficient. Given the predictions, the agent plans the signal control strategy e.g. calculating the most optimal next cycle(s). That strategy is executed by the controller agent. For specific situations, like when a large amount of traffic is expected starting T=x, with a duration of y, an agent may want to adjust its control schema. The adjusted control schema is send to adjoining agents and to the control agent. The control agent operates the traffic lights according to the outcome of the control strategy model.
During execution the control agent may extend or shorten certain green times of links to accommodate extra flow on current green links. The effectiveness of an controlled intersection largely depends on the actual traffic and the cycle plan. An optimisation procedure in which the cycle plan may be calculated minimising several criteria, such as total delay, total number of stops, total costs of losses, fuel consumption, time of day, etc., can be included in the control strategy agent.

With the conflict matrix conflict groups can be calculated. In most cases the same conflict groups are calculated. When several links are not occupied a different situation may rise. With the data of the queues the minimum green time can be calculated. Then the cycle time can be calculated. There should be specific automated help to do this kind of calculus automatically (programs like KRAAN). For real-time control the same basic program will be used. Only detectors are added to give information about queues and number of vehicles. The arrival times can also be given by the RSA so that green on demand (like green waves) could be automatically covered in the system.

3.3 System wide optimisation of the control strategy

Single junction controllers often make use of mathematical programming methodologies based on constraints for evaluation and optimisation of the cycle. For isolated traffic signals such an optimisation is already been made (Bang, 1976). For arterial and connected junctions several off-line and heuristic methods are becoming available and are still been developed but we need real-time on-line implementations (Rogier, 1999). In arterial and agent based systems this subject becomes complex due to co-ordination and synchronisation. That process in this system becomes far more complex due to several different, continuously changing, weights and different goals of the different ITSA's. Moreover, since the decisions are not centrally controlled and are based upon incomplete knowledge, the agents independently and asynchronously selects from the available feasible choices the one with the highest efficiency.

We need a provision for agents to prevent deadlocks, co-ordinate and choose between contradicting actions or non-optimal phases between agents. Metarules and authority agents will be included to handle contradictory or unexpected system wide traffic control situations. If necessary an agent can request for additional information or receive goals or orders from its authority agent. Often an ITSA should sacrifice some performance in favour of co-operative behaviour caused by appointment of an authority agent or self control of ITSA's. Some sort of negotiation is necessary in case of incompatibilities among traffic co-ordination. Negotiation then becomes a process of relaxing some of the constraints so a satisfying situation can occur; even priority alteration is possible. Also advice sought by the authority agent can be a solution for conflicts as conflicts need to be solved in real-time.
Conclusions

Signal control systems that have the capability of optimising and pro-actively adjusting the traffic light settings are able to improve vehicular throughput and delay through appropriate response current traffic. With the introduction of two un-coupled feed back loops with appropriate pro-active behaviour, whether agent technology is used or not, a different theory of traffic control can be met. Primarily results indicate that given an automated control strategy implemented in the traffic signalling devices we can get a system that makes better use of the capacity of the intersection. It has been shown that control systems based on agent technology can adapt and respond to changing conditions in real-time and thus making better use of its resources.

For urban traffic control the research agenda comprises to self adjustable control schemes that can deal with dynamic and actuated data. The particular techniques proposed are experimental and not yet mainstream, especially when proposed for such a large, on-line, application. Further research on a control strategy, based on intelligent autonomous agents, is necessary to provide appropriate evidence on the usability of AI / agent based control systems in traffic control. The pro-active and re-active nature of agents can be a helpful paradigm in intelligent traffic management and control. Further (simulated or real-life) tests on a control strategy, based on intelligent and autonomous agents, are planned to provide appropriate evidence for operational use. As this research is still ongoing we hope, in the end, to demonstrate that an integrated dynamic urban traffic control system based on agent technology can adapt and respond to real world traffic conditions in real-time. A working prototype of such a system should give appropriate evidence on the usability of AI agent based control systems.
References


- Zeigler, B.P. (1990), High autonomy systems: concepts and models, proceedings AI, simulation, and planning in high autonomy systems, Zeigler, B., Rozenblit, J (eds.), IEEE, Los Alamos
Traffic Control and Route Choice: Occurrence of Instabilities

1. Introduction
1.1 Control optimization problem
1.2 An example
1.3 Multiple solutions
2. Further analysis
3. Final remarks

TRAIRL Research School, Delft, December 1999

Authors
Ir. H. Taale and Prof. dr. H.J. van Zuylen
Department of Civil Engineering and Geosciences, Delft University of Technology
Transport Research Centre, Ministry of Transport, Public Works and Water Management

Thesis supervisor
Prof. dr. H.J. van Zuylen
Department of Civil Engineering and Geosciences, Delft University of Technology
## Contents

Abstract ................................................................. iii

1 Introduction........................................................... 1

2 Route choice based on day-to-day experience...................... 3

3 The combined traffic assignment and control optimisation problem.. 5
  3.1 Introduction....................................................... 5
  3.2 An example....................................................... 5
  3.3 Multiple solutions.............................................. 9

4 Further analysis..................................................... 12

5 Final remarks ...................................................... 17
  5.1 Conclusions ..................................................... 17
  5.2 Further research ............................................... 17

References..................................................................... 18
Abstract

Traffic control and traveller's behaviour are two processes which influence each other. The two processes have different objectives that the actors in these processes try to achieve: the road manager will try to achieve a system optimum and the road users will search for their own optimum. Decisions taken by the road manager in traffic control have an influence on the possibilities for travellers to choose their preferred mode, route and time of departure, and vice versa. Road managers and road user have different objectives. The control problem is to optimise traffic control in such a way that the system is at an optimum, taking into account the reaction of travellers.

Some simple examples were taken to study the result of the optimisation of the different actors on the dynamics of the system and the existence of stable situations where each actor has no possibility to optimise his decisions anymore. This has been done with analysis and microscopic simulation. Under certain conditions multiple stable situations are possible, but some of these situations are sensitive to small disturbances, by which the system moves away from the original situation. There appears a non-linear relationship between system parameters and the character and location of the equilibrium situations.
1 Introduction

There is a tradition in traffic control to adapt control structure and parameters to the actual traffic flows and conditions, such that delays, queues or some other collective objective function is optimised. For fixed-time and traffic actuated controllers, methods and guidelines have been developed. For example: Webster’s formula for cycle time and green splits, the rule to minimise delays by choosing green splits which minimise the maximum degree of saturation, the rules for maximum green times and maximum gap times for vehicle actuated controllers, etc.

Apart from the optimisation of the total network, often a special treatment is given to certain groups of road users. In urban areas pedestrians may get a preferred treatment; at bus routes the bus may get a priority treatment and on crossings of cycle tracks cyclist may get more frequently a green phase. Such priority control is the consequence of a policy to assign the use of traffic space especially to certain preferred groups of road users because they play an important role in the local situation. In the case of priority for public transport, reason might be that this minimises the total waiting time of all road users and improves the operating speed for busses and trams. Another reason is to reduce the operating costs of a transport mode, which is heavily subsidised by public funding.

There is also a certain expectation from the authorities that a preferential treatment of certain traffic classes - and in most cases one has public transport and pedestrians in mind - will reduce the growth of car traffic and will influence the modal choice in favour of collective public transport, cycling or walking. In each case it is possible to influence route choice and time of departure by traffic control.

Because traffic control has an influence on travel behaviour, a change in traffic control may have the impact that traffic volumes change. If traffic control is modified such that congestion on a certain route disappears and delays on intersections decreases, traffic might be attracted from other links where congestion still exists or which are part of a longer route. This might have the consequence that queues, which originally disappeared, return. Delays may come back on the original levels. The question is whether there still is a net profit for the traffic system as a whole.

If we assume that a modification in traffic control gives a change in traffic control, it is necessary to anticipate this change. If we want to optimise delays, it should be done for the traffic volumes that will be present after the introduction of the optimised traffic control and not for the traffic volumes which existed before the implementation. Of course, it is possible to follow an interactive approach, where after each shift in traffic volumes the control scheme is adjusted until equilibrium has been reached, or one may use self-adjusting traffic control.
However, it can be shown, for certain examples, that the process of the adjustment of traffic control, followed by a shift in traffic volumes, does not necessarily lead to a system optimum. It is even possible that the system oscillates between two or more states.

The objective of the paper is to show that even small problems with simple assumptions about day-to-day route choice can have multiple solutions. This can even lead to oscillations or even chaotic behaviour of the system. The complicated problem of day-to-day route choice in combination with optimisation of control is not discussed here, but is a topic for further research.

In this paper we start with the description of a simple route choice model which displays the occurrence of oscillations. In section 3 we will study a simple road network with traffic control on one of the links and we investigate the occurrence of stable, consistent equilibrium situations. First we give the results of a simulation study. This shows the more or less conventional pattern that one stable condition exists where flows are consistent with the travel times and the traffic control is optimised with respect to total delay. A further analysis of a slightly simplified situation shows that there is a possibility of the occurrence of oscillation in route choice for a simple road network. In section 4 we study the combined route choice and traffic control optimisation problem in some more detail. It is shown that in the two-level optimisation problem optimum solutions exist which can be interpreted as meta-stable saddle-points. This means that there is equilibrium between traffic control and route choice, but small disturbances in the conditions will cause the system to slide to another state. Finally, we draw some conclusions and we make some remarks on the existing problems and the directions for further research.
2 Route choice based on day-to-day experience

It is assumed that the route chosen by travellers depends on their perception of travel times on different alternative routes. Normally, travellers have imperfect knowledge about the actual travel times and have to rely on experiences in the past. It has been shown that route choice based on historical knowledge can lead to oscillating behaviour, where on one day one route is preferred and the next day most of the traffic follows another route (Nakayama et al. 1999). It is also known (Horowitz, 1984 and Watling, 1999) that a traffic assignment based on stochastic route choice does not necessarily lead to equilibrium.

A simplified model, which describes the dynamics of route choice with two alternatives parallel routes, is the following. We assume that on day $n$ the volume on route 1 is proportional to the volume on the previous day on the same route (conservative force) and also proportional to the volume on the other route (the more volume on the other route, the higher the travel times and the more travellers will switch). This model is

$$V_{n,1} = a V_{n-1,1} V_{n-1,2} = a V_{n-1,1} (V - V_{n-1,1})$$

where $V = V_1 + V_2$ and $a$ is a constant which can be determined from observations, e.g. from the equilibrium state when $V_n = V_{n-1}$. Equation (1) is an example of the logistic equation, which is known to give equilibrium states for certain values of $a$, oscillating behaviour for other (higher) values of $a$ and chaotic behaviour if $a$ comes above a certain value.

If in equation (1) the transition is made from volumes to fractions of traffic choosing a route, $x$, the equation becomes

$$x_n = a' x_{n-1} (1 - x_{n-1})$$

where $a' = aV$.

In Figures 1 and 2 the dynamics of route choice is illustrated for different values of $a'$. Transition between different dynamic patterns can be described as splitting of the number of states between which the route choice oscillates. The transition to chaotic behaviour is characterised by the fact that there is an unlimited number of states between which the route choice is moving.

The difference between random and chaotic behaviour is, that chaotic behaviour has certain regularities, such as the quasi-periodicity and (strange) attractors. Furthermore, chaos behaviour often has the property that small changes in behaviour result in large changes in the future state. The similarity is that both random and chaotic behaviour cannot be predicted. Important properties of systems with chaotic properties are:
Approach to equilibrium $\alpha' = 2.5$

Oscillating route choice $\alpha' = 3.2$

Figure 1: Route choice between two alternative routes, equilibrium and alternating behaviour

Double oscillations $\alpha' = 3.5$

Chaotic behaviour $\alpha' = 3.9$

Figure 2: Route choice: oscillations with four states and the transition to chaotic behaviour

- they have non-linear dynamics
- positive feedback exists, by which certain changes are enhanced
- negative feedback exists, which drives the condition of the system from a dynamics of unlimited growth in one direction.

These three properties can clearly be seen in the example in this section. Equation 2 is non-linear, for small values of $x$ there is positive feedback proportional to $x$ and for values of $x \approx 1$, the negative feedback pushes $x$ back in the direction of smaller values.

Of course this is a simple and maybe even an unrealistic model of route choice behaviour and much depends on the parameters $\alpha'$, but it shows that even simple examples of non-linear systems with positive feedback can behave in a very complex way.
3 The combined traffic assignment and control optimisation problem

3.1 Introduction

Already in the seventies Allsop and Charlesworth (Allsop 1974, Charlesworth, 1977) showed the relevance of the interdependence of traffic control and route choice. The problem to be solved was initially to search for a traffic control scheme that optimises total delay for traffic volumes which are consistent with the travel times influenced by the control scheme, i.e. a traffic condition where no traveller can improve his travel time by choosing another route (Wardrop’s first principle). Two parties try to achieve their own goals, each with its own objective function and space of choices. The infrastructure manager tries to optimise the road system, for example by maximising the utilisation and minimising total delays and stops. A part of the available instruments is the setting of traffic signals, but also other measures are possible. The second group of actors are the drivers who choose their routes such that they minimise their travel time. The travel times are partly determined by the traffic signal settings; while the traffic signal settings are optimised for certain traffic flows that are the consequence of the behaviour of the drivers.

Fisk (1984) showed that this situation could be seen as an example of a non-co-operative game, in which two players have their own objectives and their own strategies. The strategy is known and the choices are predictable, such that it is possible to choose an optimal strategy, taking into account the predictable reaction of the other party. Road users can choose their route under the assumption that traffic control will be optimised for the total delays. The infrastructure manager can optimise the traffic control knowing that the road users will shift their roads after the modification of the control scheme.

3.2 An example

A first attempt to analyse this problem was by simulation. The objective was to analyse the distribution of flows between origin A and destination B on two alternative routes. One route has a controlled intersection; the other one is a bypass (see figure 3).
A junction as sketched in figure 4 was chosen for this study (from a traffic network study in Hong Kong). For this junction a two-lane bypass for the north-south movement was created. This bypass is situated on the western side of the junction and is 2.5 kilometres longer than the route across the junction. The free speed for the bypass is 100 km/hr and for the route with the junction 50 km/hr. The flows for the AM-peak are as given in figure 5.
For this situation a number of control strategies were simulated to see how they affected route choice. For the calculation of travel times, the microscopic simulation model FLEXYT-II was used. The Transport Research Centre (AVV) of Rijkswaterstaat developed FLEXYT-II. A more detailed description of the model and its validity can be found in (Taale and Meuris 1995 and 1997) and (Taale and Scheerder 1998). The travel times have been calculated as the travel time at cruising speed plus the average delay at the intersection.

Because FLEXYT-II has no assignment, other than specified by the user, route choice in this study was investigated by changing the splitting rate. First, it was assumed that only 1% of the traffic took the bypass (to measure travel time), then the splitting rate was increased to 10%, 20%, etc. In this way it was possible to find the equilibrium. This was done for four control types. First, the existing fixed-time control plan was simulated for all splitting rates. Then the same was done for optimised fixed-time control. Using Webster’s formula for cycle time and green times, the optimised control plan was derived for the original flows. Then, the existing vehicle actuated control plan was used and finally an optimised vehicle actuated control. This was done also by using Webster’s formula and putting the green times as maximum green times in the control plan.

In figure 6 the travel times on both routes are shown for all control types. The travel time for the bypass is the same for all types.

**Figure 5: Traffic flows, AM-peak**
Figure 6: Travel times for different distributions of the flow on the controlled route and the bypass

For normal fixed-time control equilibrium is never reached. For optimised fixed-time control the equilibrium is around 50%. For both vehicle actuated control types equilibrium is reached around 60%. The total delay is shown in figure 7.

Figure 7: Total delay in the network
The figure shows that a user optimum does not necessarily mean a system optimum. From the previous figure it was clear that a user optimum was located around 50% or 60%, but from figure 7 it can be seen that a system optimum is reached when as much traffic as possible uses the bypass. So far, the green times were optimised only for the situation in which only 1% of the traffic uses the bypass. When the optimisation is done for both equilibrium situations, the following table is the result.

<table>
<thead>
<tr>
<th></th>
<th>Fixed-time (splitting rate 50%)</th>
<th>Vehicle actuated (splitting rate 60%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>before</td>
<td>after</td>
</tr>
<tr>
<td>total distance travelled</td>
<td>10888.97</td>
<td>10823.72</td>
</tr>
<tr>
<td>travel time bypass (sec.)</td>
<td>225.9</td>
<td>225.8</td>
</tr>
<tr>
<td>travel time junction (sec.)</td>
<td>224.8</td>
<td>224.9</td>
</tr>
<tr>
<td>total delay (veh.hrs/hr)</td>
<td>47.91</td>
<td>39.14</td>
</tr>
</tbody>
</table>

Table 1 shows that for fixed-time control the travel time for the two separate routes does not change and that the total delay decreases with 18%. For vehicle actuated control the situation does not change at all: the travel times on the two routes and the total delay stays the same.

3.3 Multiple solutions

The result of the simulation study showed one single solution for the combined assignment and traffic control problem. Multiple solutions were obtained for a similar problem as figure 3:

![Figure 8: Example of a network with symmetric choice possibilities](image)

Figure 8: Example of a network with symmetric choice possibilities
The example of figure 8 is symmetric. Both routes between A and B cross the route between C and D at a controlled intersection. The traffic control at these intersections is fixed time optimised with Webster's method. It appears that, depending on the magnitude of the flows and the internal lost time, several different solutions exist:

- the symmetrical solution with a 50 - 50% distribution between route 1 and 2
- an asymmetrical solution where more drivers choose for one of the two routes.

Of course, for every asymmetrical solution a 'mirror' solution exists: if a stable solution is obtained with x% on route 1 and 100 - x% on route 2, another solution is the distribution 100 - x% on route 1 and x% on route 2.

In figure 9 the difference in travel time between route 1 and 2 is given for different values of the percentage of the traffic, which is using route 1. The travel time difference is the same as the difference in delay and that was calculated using the two-term delay formula derived by Webster (Webster, 1958, see also chapter 4).

The stable situations occur if the difference in travel time is zero. Figure 9 gives the symmetrical situation where the equilibrium and optimum is obtained for 50% distribution:

![Figure 9: Travel time difference between route 1 and 2, giving one single symmetric equilibrium](image)

Traffic Control and Route Choice: Occurrence of Instabilities
The situation was calculated for 1000 veh/h flow from A to B, 500 veh/h from C to D, saturation flows of 1800 veh/h and internal lost times of 9 seconds, with a minimum green time of 6 seconds. If we change the flow from A to B to 600 veh/h, the picture changes a lot:

![Travel Time Difference Graph](image)

Figure 10: Travel times differences for route 1 and 2 with asymmetric equilibrium

Two asymmetric stable equilibrium states exist: the 20 - 80% and the 50 - 50% distribution. If we have the system in the 50-50% state and small changes occur in this distribution, the change is enhanced by the subsequent adaptation of the traffic control scheme: the control scheme for the route with the largest flow gives shorter average delays. The total travel time is in both cases at a minimum for the 50 - 50% distribution. Changing the parameters of the control scheme (lost time or minimum times) or changing the flow or saturation flows, changes the appearance of the time-difference curves significantly, so that a small change of the parameters can have the result that the equilibrium states move over large distances and the asymmetric solution disappears suddenly. Apparently the system of route choice and traffic control is under certain circumstances critically dependent on system parameters. In the following chapter we shall investigate this behaviour in some more detail for the original network of figure 3.
4 Further analysis

The assignment problem in the case of deterministic route choice based on individual shortest routes can be formulated mathematically as

$$\min_{V_i} Z(V_i)$$

(3)

where $V_i$ is the volume on link $i$. The function $Z$ is defined as

$$Z = \sum_i \int T_i(z,C,t_g) \ dz$$

(4)

and $T_i$ is the traveltime on link $i$ including delays for volume $z$, cycle time $C$ and greentime $t_g$. The minimisation of the delays on controlled intersections can be represented by the following formal expression

$$\min_{t_i} \sum_i D_i(V_i, t_i)$$

(5)

where $t_i$ are the time parameters of the traffic control, $D_i$ is the delay for link $i$ and $V_i$ represents the volumes to be calculated from the solution of equation (3).

In order to get some more insight in the characteristics of the problem, we shall reduce the traffic control problem to one single dimension. The combined assignment and optimisation problem can be visualised in a two dimensional space which makes further analysis easier. We assume that the cycle time remains fixed, the only parameter left is the green split. The delay $d$ for a single controlled flow is given by (Webster, 1958):

$$d(V, C, t_g) = 0.9 \left[ \frac{1}{2} (C - t_g)^2 (1 - V / s)^{-1} C^{-1} + \frac{1}{2} x^2 / V (1 - x) \right]$$

(6)

with

- $V = \text{volume}$
- $x = (V / s) (C / t_g)$
- $s = \text{saturation flow}$
- $C = \text{cycle time}$
- $t_g = \text{green time}$.

For both approaches of the intersection together the delay is given by

$$D \approx 0.9[V_1\left(\frac{1}{2}(C-t_{g1})^2(1-V_1/s_1)^{-1}C^{-1}+\frac{1}{2}(V_1C/t_{g1}s_1)^2/V_1(1-V_1C/t_{g1}s_1)\right)+V_0\left(\frac{1}{2}(C-t_{g0})^2(1-V_0/s_0)^{-1}C^{-1}+\frac{1}{2}(V_0C/t_{g0}s_0)^2/V_0(1-V_0C/t_{g0}s_0)\right)]$$

(7)

with:

- $0 \leq V_1 \leq V$
- $t_{g1} + t_{g0} = C - t_i$ (ti is the internal lost time of the control scheme).
Figure 11: Total delay as a function of the green time $t_g$ and volume $V_I$.

Figure 12: Iso-curves with equal total delay in the $t_g-V_I$ plane. The thick line gives the green time that minimises the total delay for a given $V_I$. 

TRAIL Research School, December 1999
The route choice problem can be formulated in the following way

\[ \min_{(V_1, V_2)} Z(V_1, V_2) \text{ with } V_1 + V_2 = V \]  \hspace{1cm} (8)

where \( Z \) can be elaborated into

\[
Z = V_1L_1/v + V_2L_2/v + 0.45[(C-t_g)/C \int (1-z/s)dz + (C/s.t_g)\int z(1-z/C/s.t_g)dz] \]

\[
= V_2L_2/v + V_1L_1/v + 0.45[s(C-t_g)^2/C \ln(1-V_1/C/s.t_g)-V_1C/s.t_g\ln(1-V_1/C/s.t_g)] \] \hspace{1cm} (9)

With the boundary condition \( V_1 + V_2 = V \), the function \( Z \) becomes also a function of two variables, \( V_1 \) and \( t_g \).

![Objective function assignment](image)

**Figure 13:** Objective function \( Z \) as a function of \( t_g \) and \( V_1 \).

Figure 13 gives a graphical representation of \( Z(V_1, t_g) \). The equilibrium solutions are on the line drawn in figure 14 where \( \delta Z / \delta V_1 = 0 \).

If we combine the lines, which give the optimum green split (from figure 12) and the equilibrium assignment (figure 14), we obtain figure 15:
Figure 14: Iso-lines with equal values of the objective function $Z$ and the (thick) line giving the equilibrium assignment for a given green time.

Figure 15: Optimum green time (dotted line) and equilibrium assignment (drawn line) with three equilibrium situations.
We see that in this example three situations exist where the traffic control is optimized with respect to the traffic volumes and the traffic volumes are consistent with the travel times. If we assume that the process of adjustment of traffic control and route choice are iterative, we find patterns given by the arrows in figure 15: an adjustment in traffic control will give a change in travel time, with the consequence that some drivers choose another route. The changed traffic volumes make it necessary to adjust the control scheme etc. The process stops if a situation has been reached where the drawn and dotted curves intersect (i.e. points 1, 2 and 3).

If in situation 2 the route choice would slightly change and the traffic control is adapted to the changed flows, we see that a positive feedback mechanism exists: a small variation in route choice is reinforced by the mechanism in which more traffic leads to more green time which reduces delay and attracts more traffic, etc. Only at the extremes, where all traffic chooses the same routes or where congestion prevents further growth, the positive feed will disappear. So in this example two stable situations exist: 1 and 3. The total travel time is minimal (for this example) in situation 3.

Also in this case the form of the two curves of figure 15 depends critically on control parameters and (saturation) flows. The optimal green split depend on minimum green time and the internal lost time, which makes that the shape of the curves in figure 15 is determined for a great deal by the boundaries of the space of feasible solutions. Changes in the boundaries will change the shape of the curves, which can have the consequence that the curves intersect on one, two or more points and that the intersection point can move irregularly after small changes in the system parameters or boundary conditions.

The increasing and decreasing difference between the curves of the optimal green split and the equilibrium assignment is due to the non-linear behaviour of the delay function. Small increases in green time lead to large decreases in delay and thus a large switch in traffic from route 2 to route 1. This effect is larger than the increase in delay due to the increase in traffic flow. But at a certain level the increase in traffic will compensate the decrease in delay caused by the larger green times and the difference between the curves will be smaller.
5 Final remarks

5.1 Conclusions

In a few simulation studies and a further analysis of the problem it is shown that in rather simple traffic situations very complex processes can arise, if we let the system move to equilibrium. The equilibrium situation is not always uniquely determined and it is even possible that oscillations occur (Chen and Wang 1999). The equilibrium situation that is achieved after an iterative adjustment of traffic control to changing route choice is not always a system optimum. This leads to the conclusion that the traffic dependent optimisation of traffic control may result in a sub-optimal situation and it might be better to use traffic control as a management tool to steer the traffic flows, more than as a mean to accommodate traffic volumes.

5.2 Further research

The necessary analytical tools for such a strategic approach are still limited: the combination of dynamic traffic control and dynamic traffic assignment that supports the search for a system optimum is still to be developed and is an important topic for further research. Furthermore, the knowledge about the occurrence of the instabilities is very limited yet. Empirical data on this subject may exist; there are even real time systems that estimate travel behaviour from real time traffic data (e.g. Bell and Grosso 1998). However, as far as the authors know, no analysis of the existing traffic data has been reported which looks for the existence of multiple stable equilibria or to the possibilities to increase system performance by changing traffic control and route choice simultaneously.

The problem becomes more complex even if we realise that in the real world the degree of freedom for travellers is much larger than just route choice. The influence of traffic control on time of departure, modal choice, frequency of travelling, choice of destination etc. has been quantified by a few researchers (Mokhtarian and Raney 1997, Mogridge 1997), but more should be done to apply these results to a method which optimises traffic control taking into account the expected behavioural response, including day-to-day dynamics. Apart from practical tools, which make it possible to optimise traffic control, predict the impact on travel behaviour and to anticipate on future change in behaviour, there is also a need for an analytical framework to study the existence of equilibrium conditions in a system of traffic control and individual travellers.
References


References


Drivers' Stated Preferences for a Driver State Monitoring System

TRAIL Research School, Delft, December, 1999

Authors
Ir. V.A.W.J. Marchau, Dr. E.J.E. Molin and Drs. S. van Hoytema
Faculty of Technology, Policy and Management, Delft University of Technology

Ir. D. de Waard
Centre for Environmental and Traffic Psychology, University of Groningen

Thesis supervisor
Prof. dr. ir. R.E.C.M. van der Heijden
Faculty of Technology, Policy and Management, Delft University of Technology
Contents

Abstract

1 INTRODUCTION 1

2 THE SAVE-SYSTEM 3

3 CONJOINT ANALYSIS 5
  3.1 The concept 5
  3.2 Selection of attributes 6
  3.3 Determination of attribute levels 6
  3.4 Selection of experimental design 7
  3.5 Choice of measurement task 8
  3.6 Choice of estimation procedure 8

4 METHODOLOGY 9

5 RESULTS 12
  5.1 Response group characteristics 12
  5.2 The overall preference model 13
  5.3 Attribute importance 15
  5.4 Predictive validity of completely disaggregate models 16

6 CONCLUSION 17

References 19
Abstract

The introduction of new electronic driver monitoring systems in road vehicles is expected to improve traffic safety significantly. These systems are intended to guard the alertness of drivers and, if necessary, warn the driver and/or take over control to stop the car in a safe way. Successful implementation of these systems will, among others, depend on the willingness of people to buy and use these systems. In this paper, the acceptance of users is explored regarding the SAVE system (System for effective Assessment of driver state and Vehicle control in Emergency situations). Attractiveness of SAVE has been measured using a decompositional stated preference approach. A group of 78 potential users have been questioned about their preferences regarding several alternative SAVE options. This paper presents and discusses the estimated part-worth utilities of six functional and cost related attributes.

keywords: driver state monitoring system, consumers' stated preferences
1 Introduction

Throughout the last decades the number of traffic accidents steadily increased, causing on average about 1,250,000 injured and about 45,000 deaths in European Union each year in the nineties (European Commission, 1999). This number has inspired the European Commission to finance research and development programmes on technologies to diminish the accident rates. Part of this research and development focus on systems which intend to improve road vehicle control. In general, the area of road vehicle control is concerned with the application of vehicle-technologies which assist the driver in performing his driving task in a better way by partially or even totally automation of these driving tasks. One of the projects financed by this program is the SAVE project (System for effective Assessment of driver state and Vehicle control in Emergency Situations). The aim of this project is to develop an electronic system that will monitor driver state in real time and, if judged necessary, undertake emergency handling, prior and during the emergency situation. SAVE potentially has a large effect on road safety as it can detect and act upon important causes of accidents. Smiley and Brookhuis (1987) concluded that some 90% of all traffic accidents can be attributed to human failure in general and it is estimated that at least 30% of all serious car accidents must be attributed to problems concerning driving state (e.g. Hayward et al., 1997; Bassuge, 1995), such as alcohol or drug abuse, fatigue, inattention and health problems (e.g. heart attack, epileptic seizure, fainting). A potential positive effect of SAVE will naturally be larger if more car drivers purchase and use the system. Currently there is limited knowledge regarding these issues. Insight in the drivers’ preferences is therefore necessary to further develop the system in a direction that will maximise market penetration.

When driver’s preferences or attitudes towards new technology are studied (e.g. Brookhuis and Soeteman, 1998; Hoedemaeker, 1996; Hoedemaeker et al., 1996; Bekiaris et al., 1997), respondents usually are requested to respond to items that each describe one attribute of the technology to be studied. Respondents are presented each attribute in turn and are requested to explicitly express on some measurement scale to what extent they desire the attribute, they feel that the attribute is important or evaluate the attribute in some other way. Questionnaires including this measurement method are relatively easy to construct and fairly easy for respondents to complete. Moreover, the responses potentially have high reliability, that is, when the questionnaire is repeated after a certain time period with the same respondents, practically the same results are obtained. However, doubts have been risen in the literature about the validity of such measurement approach: does the instrument measure what it is supposed to measure?
Comparative research in the marketing and spatial science literature has repeatedly reported that respondents have trouble in explicitly stating their preferences for attributes they have to evaluate separately. The importance of less important attributes tends to be overestimated while the importance of truly important attributes tends to be underestimated. The probable reason for these findings is that while usually more than one attribute plays a role in the decision making process and therefore decision makers have to make trade-offs between attributes to arrive at overall preferences, this is largely neglected by the traditional measurement techniques.

The trade-off between attributes is explicitly taken into account by an alternative measurement approach called the decompositional stated preference approach, also called conjoint analysis (Louviere, 1988). This approach makes use of hypothetical profiles which integrally describe products or services and to which respondents are requested to provide overall evaluations or to make choices. Because the profiles are constructed according to the principles of statistical designs, the effect of each attribute on the overall preference or choice can be estimated efficiently and without bias. Providing overall preferences or choices for hypothetical profiles is probably more difficult for respondents than separately evaluating each attribute, which potentially leads to less reliable results. However, as complete products or services are regarded with the result that the measurement task bears more resemblance with the behaviour in real markets and the trade-off between the attributes is taken into account, the estimated preferences potentially have higher validity. In general, methods are preferred that provide results with higher validity to methods that provide less valid results, even if these results are less reliable. Therefore, the decompositional stated preference approach is applied in this research project to estimate the car drivers’ preferences for the SAVE system. The results of this research project are presented in this paper.

This paper is organised as follows. First, the components of the SAVE system will be briefly described. Next, for readers who are not familiar with the decompositional stated preference approach, this approach will be shortly explained. This is followed by an outline of the methodology applied in this research project. Then, the estimated preference model will be presented and discussed. This paper finishes with drawing some conclusions.
The SAVE system is an in-vehicle device that in real-time detects deterioration in driver performance as a result of alcohol abuse, fatigue, breakdown or inattention. The components of this system are briefly described in this section.

The SAVE system exists of five functionally different subsystems (Brookhuis et al, 1998): the Hierarchical Manager (HM), the Integrated Monitoring Unit (IMU), the Human Machine Interface (HMI), the SAVE Warning System (SWS), and the Automatic Control Device (ACD). Figure 1 shows the relationships among these subsystems and between these and the driver, the vehicle and the environment.

The Hierarchical Manager (HM) is the central processing unit, the ‘brain’ of the SAVE-system. Its principal function is control over the subsystems. HM performs diagnoses, integration, prioritisation and determination of actions in real time.

The Integrated Monitoring Unit (IMU) primary objective is to continuously monitor the driver’s ability to drive safely, and can be considered the ‘eyes’ of the system. The IMU collects data from various sensors that monitor the driver, the vehicle and the environment. Examples of data collected from the driver include eye-lid closure, head position, and grip force on the steering wheel. Relevant vehicle data are speed, steering wheel angle, distance to leading vehicle, lateral position, and time to line crossing. Data concerning the environment are time of day, weather conditions and environment (e.g., city or free way). By using these sensor data, SAVE can distinguish between a driver’s normal and abnormal driving behaviour. SAVE first has to learn the normal driving behaviour while the driver drives under normal circumstances. By comparing the actual behaviour with the normal behaviour, the systems can detect abnormal behaviour if critical values have been exceeded. SAVE uses both absolute and relative critical values. Absolute critical values apply for every driver and point to acute danger, for instance, when the time-headway to a lead car is smaller than the minimal human reaction time. Relative critical values are individual and refer to the worsening driving behaviour due to driver’s inattention or malfunctioning. For example, the car’s headway to a lead car is smaller than usual. All measured sensor values and critical values are stored in a database.

The function of the SAVE-Warning System (SWS) is to warn the driver if the HM concludes that the driver does not behave safely, but still is able to stop the vehicle. If critical values have been seriously exceeded and the driver does not respond to warnings, SWS will also warn other road users and a traffic control centre.
The Human Machine Interface (HMI) serves as the communication link between the driver and the system. On the one hand, HMI provides information and warnings to the driver. On the other hand, the driver provides information to the system by way of using HMI and hence, controls the system to a certain extent. A smart card reader feeds personal driver information into the system, which in the future will be extended with an anti-theft device through start disabling.

Finally, the Automatic Control Device (ACD) safely stops the car automatically along the roadside if, and only if, HM diagnoses that driving safely is unacceptably jeopardised. This is the case only when no response is detected after a warning has been issued by the SWS, or if the situation is considered acutely critical. Once started, ACD operates independently from other subsystems, although it makes use of the data from existing sensors and actuator. This enables ACD while stopping the car to take account of the riding-tracks and the surrounding traffic.

Once fully developed, SAVE can correctly detect the following causes of worsening driving behaviour: acute serious illness (heart or epileptic attack, fainting), fatigue or drowsiness, and alcohol or drug abuse. An experiment in a driving simulator indicated the 84% to 97% of the inattention cases could be detected properly (De Waard et al., in press). These figures indicate that although SAVE may not yet be sufficiently reliable for market introduction, it certainly has potential.
3 CONJOINT ANALYSIS

For the reader who is not familiar with conjoint analysis, this section introduces this modelling approach.

3.1 The concept

Conjoint analysis is based on a concept of choice behaviour as presented below. Choice behaviour is considered to be the outcome of an individual decision-making process. The outcome is an individual's act of choosing a particular alternative from a set of potential alternatives under consideration. Each choice alternative is characterised by a large but finite number of objective attributes. Individuals are assumed to evaluate the values or levels of these attributes and to combine these according to some combination rule which they use to form an overall evaluation of each alternative under consideration. This cognitive process involves a subjective weighing based on the individuals' personal information system, which they have gathered through search and learning processes and is related to their value system, motivation and possibly to other more objective personal characteristics. The result of this process is the formation of a subjective preference scale. A preference scale may be conceived of as some composite of the subjectively weighted attribute levels, where the weights indicate the relative importance an individual assigns to the attribute levels. It consists of an ordering of the choice alternatives based on their utility in satisfying their particular needs. The actual choice behaviour of an individual is, then, to decide which alternative to choose from his evoked set. A reasonable assumption is that overt choice behaviour bears some systematic relation with the positioning of the choice alternatives on the preference scale. Choice behaviour then involves an implementation of a decision rule. It is often assumed that the individual will choose the alternative with the highest preference scale value.

The objective of conjoint analysis is to estimate preference functions, which describe how individuals combine their evaluations of attribute levels that make up the choice alternatives to arrive at overall preferences. To that effect, conjoint analysis makes use of hypothetical profiles to which respondents are requested to express their overall preferences or make choices. These profiles are integral descriptions of relevant attributes and are constructed according to the principles underlying the design of statistical experiments. The respondents' responses to these profiles are used to estimate preference or choice models. In order to arrive at such models, the following steps are often distinguished (Molin, 1999):

TRAIL Research School, December 1999
i. selection of attributes
ii. determination of attribute levels
iii. selection of experimental design
iv. choice of measurement task
v. choice of estimation procedure.
Let us consider each of these steps in turn.

3.2 Selection of attributes

The first step in constructing a conjoint experiment concerns the selection of the most salient attributes that influence the choice behaviour of interest. Important attributes cannot be omitted, because if they are, respondents may infer their values from the included attributes which will result in biased effect estimates of the included attributes. Furthermore, if the aim of the model is to inform managers or policy makers about the consequences of their actions, at least part of the attributes should be controlled by them. The selected attributes should then be retained, combined, or reformulated to keep their number as small as possible to make an experiment tractable. Naturally, the attributes have to be expressed in a way that they connect with the respondents' cognitive representations of choice alternatives.

3.3 Determination of attribute levels

Once attributes are selected, one has to decide which levels to use to form profiles. Usually, two to four levels are distinguished. For categorical attributes, the number or categories may be determined by the degree of detail required to support managerial decisions. For continuous attributes, the number of levels is dependent on the assumptions one is willing to make about the relationship between the attribute values and the derived utility. If one assumes that the part-worth utility linearly increases or decreases with increasing attribute values, only two attribute values are required. If one assumes that the utility function is better be represented by a curve, one needs to select at least three levels. Finally, if the relationship is assumed to be more complex, one needs to select at least four values. To ease the interpretation of the results, the values are usually chosen to have equal intervals. The range of attribute levels is chosen such that they span the range observed in current or intended choice alternatives. This principle is based on the idea that while it is valid to interpolate preference values for attribute levels within the range of selected values, the validity of the part-worth utility of extrapolated attribute values may be of concern.
3.4 Selection of experimental design

Once attribute levels have been selected, they have to be combined to construct full profiles that can be presented to respondents. The number and composition of profiles is dictated by the choice of an experimental design. In turn, the latter is determined by one's assumptions about the strategy decision-makers use to integrate their part-worth utilities into overall utilities for choice alternatives. Often it is assumed that decision-makers simply add their part-worth utilities to arrive at overall utilities. An additive model may represent this combination rule, in which two types of effects can be distinguished: main effects and interaction effects. A main effect is the part-worth utility contribution of an attribute-level to the overall utility regardless of the occurrence of other attribute-levels in the choice alternative. An interaction effect is the contribution to the overall utility caused by the specific combination of two or more attribute levels which is significantly higher or lower than the mere sum of the corresponding main effects.

In order to choose an appropriate experimental design one has to trade-off the possibility to estimate and test interaction effects against the requirement to reduce the number of profiles in order to avoid respondent fatigue and reduce the complexity and expense of the data collection effort. Full factorial designs allow one to estimate and test all possible interaction effects. This design involves all possible combinations that can be made of the selected attribute levels. As it is clear that if the total number of selected attribute levels is high, the number of combinations easily becomes too large to present to individual respondents. To reduce the number of profiles, one often chooses fractional factorial designs, which usually involve an orthogonal fraction of the full factorial design, examples of which can be found in Steenkamp (1985) and Addelman (1962). To apply fractional factorial designs one must assume that certain interaction effects are not statistically significant. However, often the main effects in these designs are not independent of these assumed zero interactions. If this is the case and interactions are not equal to zero (although this cannot be tested), the estimated main-effect are biased. If one assumes that all interaction effects are not significant, one arrives at main-effects-only designs, which most strongly reduce the number of profiles constructed.
3.5 Choice of measurement task

After the profiles have been constructed, one must decide which kind of response is requested from the respondents. One can distinguish ranking, rating and choice tasks. A ranking task involves respondents to rank-order the set of profiles with respect to their overall preference. Rating tasks involve that each profile is evaluated on some rating scale. A choice task involves respondents to choose between two or more profiles. The construction of choice tasks involves an additional step, as the profiles have to be placed in choice sets (e.g. Oppewal and Timmermans, 1992). One way of choosing between the distinguished measurement tasks is to consider the choice behaviour in real markets and reflect this in the response task as much as possible.

3.6 Choice of estimation procedure

Dependent on the type of data generated by the measurement task, an appropriate estimation procedure is selected to estimate the preference function. Ranking data have to be analysed with nonmetric techniques. If rating data are collected, Ordinary Least Square (OLS) regression techniques are commonly applied to estimate the preference functions. Finally, if discrete choices between two or more choice alternatives are collected, maximum likelihood techniques are commonly used to estimate the parameters.

In order to estimate the effects of categorical attributes and standardise the levels across the attributes, the attributes may be coded. To that effect, different coding schemes can be applied of which effect coding is discussed here as an example. Table 1 shows that the L levels of an attribute are coded with L-1 indicator variables. The parameters estimated for these indicator variables can be applied to derive the part-worth utilities for the attribute levels by multiplying the parameters with the coded values and summing across the indicator variables. By applying regression analysis with effect coded attributes, the regression intercept is equal to the mean observed profile ratings, while the attribute level part-worth utilities are expressed in deviation from this mean. Note that the sum of an attribute's part-worth utilities is equal to zero by definition.

Table 1 Effect coding for three-level attributes

<table>
<thead>
<tr>
<th>attribute-level</th>
<th>first indicator variable</th>
<th>second indicator variable</th>
<th>derived part-worth utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>$\beta_1$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>$\beta_2$</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-(\beta_1+\beta_2)</td>
</tr>
</tbody>
</table>

Drivers' Stated Preferences for the SAVE Systems
4 METHODOLOGY

This section describes the methodology applied in this research project. The five steps just described to arrive at conjoint models will be followed. A list of potential interesting attributes was selected from a literature search (Bekiaris et al., 1996; Bekiaris et al., 1997; Bekiaris et al., 1998; Brookhuis et al., 1998; Hartley, 1995; Hawkins, 1992; Heijer et al., 1998; Wilkie, 1986) and some interviews with experts and potential users. This resulted in a list of 19 potentially relevant attributes. It is clear that this number of attributes is too large to perform our research in a proper way. Conjoint analysis usually limits to about 8 attributes, to gain reliable and valid models. Hence, the potential attributes have been analysed in-depth in order to select the most determinant attributes (Van Hoytema, 1999). This resulted in a list of six salient attributes. The six attributes selected to be included in this study listed in Table 2 show that there is an emphasis on functional and cost-related attributes.

Table 2 Selected attributes and their levels

<table>
<thead>
<tr>
<th>1 detection (when?)</th>
<th>4 additional functionality (ISA)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>- acute illness</td>
<td>- no Intelligent Speed Adapter (ISA)</td>
</tr>
<tr>
<td>- acute illness + drowsiness</td>
<td>- ISA warning</td>
</tr>
<tr>
<td>- acute illness + drowsiness + drug- &amp; alcohol abuse</td>
<td>- ISA limiting</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 control (how?)</th>
<th>5 costs (purchase costs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- warning only</td>
<td>- NlG.0 (for free)</td>
</tr>
<tr>
<td>- take over control after not reacting to warning</td>
<td>- NlG.1500</td>
</tr>
<tr>
<td>- directly taking over control</td>
<td>- NlG.3000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3 type of roads (where?)</th>
<th>6 insurance premium reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>- free ways only</td>
<td>- 0%</td>
</tr>
<tr>
<td>- all road types outside built-up area</td>
<td>- 25%</td>
</tr>
<tr>
<td>- all roads (outside and inside built-up area)</td>
<td>- 50%</td>
</tr>
</tbody>
</table>

The 'detection' attribute indicates which of the three major forms of driver impairment SAVE is able to detect. In a cumulating order: acute illness, like fainting, losing consciousness or heart attack, drowsiness, and drug or alcohol abuse. While it is expected that drivers will positively evaluate SAVE's detection for acute illness and drowsiness, the evaluation for the detection of drug and alcohol abuse is not clear. The 'control' attribute specifies how SAVE will intervene. Previous research indicated that this is a crucial factor in the SAVE acceptance. When critical values have been exceeded, SAVE can either (i) only warn the driver that something is wrong, it can (ii) warn the driver and after he does not respond take over control to safely stop the vehicle, or it can (iii) directly take over control without warning first.
The third attribute, ‘type of roads’, points to where SAVE will function: on motorways only, on all road types outside the built-up area, or on all road types inside and outside the built-up area. The next attribute, ‘ISA’ (Intelligent Speed Adapter) is included to examine whether SAVE preference can be increased if an additional functionality is added for the same price. ISA is an in-vehicle equipment that reacts to the local speed limit and warns the driver if speed limits are exceeded or automatically adjusts the car’s speed to this limit (e.g., Molin and Timmermans, 1998 and Brookhuis and De Waard, 1996). ISA can either not be combined with SAVE, it can warn only if the current speed limit is exceeded, or it can automatically limit the vehicle to the maximum allowed speed. Finally, the last two attributes are related to costs. First, ‘purchase costs’, which varies from the level ‘for free’ to Nlg.3000 (about Euro 1360). Second, a decrease of ‘operational costs’ in the form of a reduction of the insurance premium as a result of a drop in accident chance due to SAVE. The levels vary from ‘zero’ (no reduction), to 50% reduction.

The selected attribute-levels then have been combined to form experimental profiles. To maximally reduce the number of profiles a main-effects-only design was chosen to construct the profiles. An orthogonal fraction of the full-factorial design involving 18 profiles was selected. In addition, in order to test the predictive validity of the estimated model, 2 holdout profiles have been included. Holdout profiles are additional profiles that are evaluated by the respondents but are not used to estimate the model.

Respondents were requested to rate each profile on a 11-point rating scale ranging from 0 very unattractive to 10 very attractive. In addition, respondents had to indicate whether or not they would purchase the system when they would buy a new car. However, to limit the focus of this paper, this measurement will not be analysed in this paper. An example of the complete measurement task is provided in Figure 2.

![Figure 2: Example of the measurement task](image)
In order to avoid order effects, the experimental profiles were placed in three different orders resulting in three different questionnaires. Before the experiment, attributes and attribute levels are extensively explained. After introduction of each attribute, respondent ranked the attribute levels to break down the rather long explanation and in order to improve the respondents' understanding of the meaning of the attribute levels. After explaining the measurement task, respondents completed two trail profiles. After the experiment, respondents completed some personal questions about age, sex, educational level, car ownership, annual mileage (in kilometres) and whether they had health problems that effected their driving.

Potential respondents were approached in different ways. Private car drivers were approached in public places were they had to spend some time waiting, like road restaurant, cafes, hospital waiting rooms, in trains, and at sport clubs. Professional drivers, like lorry or taxi drivers, were approached at garages, driving schools and road restaurants. Most of the respondents directly completed the questionnaire at location, while others took it home and returned it by mail.
5 RESULTS

This section first presents some response group characteristics. Then, the estimated preference model is presented. This is followed by a discussion of attribute importance and predictive validity.

5.1 Response group characteristics

In the spring of 1999 in total 100 questionnaires were handed out, of which 78 usable questionnaires were returned. Table 3 presents some characteristics of the response group. From this table, it becomes clear that each category is represented by a reasonable number of cases. Hence, as each category is represented by a sufficient number of cases, it is justified to consider the measured preferences as a first insight in how drivers feel about the SAVE system.

Some noticeable observations on the response groups characteristics include the following: (i) males responded more than females; (ii) more than half of the respondents belonged to the 25-39 age group, while only a few respondents belong to the youngest group of drivers; (iii) although the categories of yearly driven kilometres are about equally filled, the drivers that drive most responded relatively more; (iv) while most of the educational level categories responded about equally, the higher vocational education category responded far more often; (v) almost all respondents own a private car, a small number has a lease car and an equal small number sometimes rents a car; (vi) finally, none of the respondents reported health problems that effect their driving behaviour.

Table 3 Response group characteristics (N=78)

<table>
<thead>
<tr>
<th>SEX</th>
<th>male 61.5%</th>
<th>female 38.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>18-24 6.4%</td>
<td>25-39 56.4%</td>
</tr>
<tr>
<td></td>
<td>40-59 25.6%</td>
<td>60+ 11.5%</td>
</tr>
<tr>
<td>CAR OWNERSHIP</td>
<td>private 80.7%</td>
<td>lease 9.0%</td>
</tr>
<tr>
<td></td>
<td>sometimes rent 9.0%</td>
<td>none 1.3%</td>
</tr>
<tr>
<td>ANNUAL MILEAGE (IN KM)</td>
<td>0-5.000 19.2%</td>
<td>5-10.000 10.3%</td>
</tr>
<tr>
<td></td>
<td>10-15.000 12.8%</td>
<td>15-20.000 12.8%</td>
</tr>
<tr>
<td></td>
<td>20-25.000 17.9%</td>
<td>25.000+ 26.9%</td>
</tr>
<tr>
<td>HEALTH EFFECT: DRIVING</td>
<td>effecting 0.0%</td>
<td>not effecting 100.0%</td>
</tr>
</tbody>
</table>

Drivers' Stated Preferences for the SAVE Systems
5.2 The overall preference model

The observed ratings for the SAVE systems profiles data were analysed by applying OLS regression analysis. The $R^2$ is equal to 0.138, which indicates that this model does not fit the individual data very well. This suggests that preferences differ considerably between the respondents. Table 4 presents the estimated part-worth utilities of the attribute levels. As effect-coding was applied to code the attribute levels, the part-worth utilities can be interpreted as the attribute level's contribution to a system's total utility expressed as deviation from the overall utility.

Table 4 shows that the overall utility respondents derived from 18 SAVE profiles is equal to 5.36. As the systems have been rated on a 0-10 point scale, this is slightly higher than the scale middle. From this result may be concluded that on average the respondents are not unfavourably but also not favourable towards the presented SAVE systems.
Let us first consider the part-worth utilities derived from the attribute that indicate when, how and where the SAVE system intervenes. With respect to 'when' SAVE intervenes, the part-worth's of the 'detection' attribute indicate that utility increases with increasing number driver impairment causes SAVE is able to detect. With respect to 'how' SAVE intervenes, the 'control' attribute indicates that drivers prefer that they are warned before SAVE takes over control. Systems that automatically take over control are disliked, while systems that only warn take an intermediate position. With respect to 'where' SAVE intervenes, the 'type of roads' attribute points out that drivers prefer that SAVE functions on all road types. Furthermore, drivers derive a higher utility from systems functioning on motorways only than those functioning on all road types outside built-up areas. This is contra intuitive as the latter category also includes motorways. Probably a

<table>
<thead>
<tr>
<th>Table 4 Estimated part-worth utility contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>part-worth</td>
</tr>
<tr>
<td>utility</td>
</tr>
<tr>
<td>1 detection (when?)</td>
</tr>
<tr>
<td>- acute illness</td>
</tr>
<tr>
<td>- acute illness + drowsiness</td>
</tr>
<tr>
<td>- acute illness + drowsiness + drug- &amp;</td>
</tr>
<tr>
<td>alcohol abuse</td>
</tr>
<tr>
<td>2 control (how?)</td>
</tr>
<tr>
<td>- warning only</td>
</tr>
<tr>
<td>- take over control after not reacting to</td>
</tr>
<tr>
<td>warning</td>
</tr>
<tr>
<td>- directly taking over control</td>
</tr>
<tr>
<td>3 type of roads (where?)</td>
</tr>
<tr>
<td>- free ways only</td>
</tr>
<tr>
<td>- all road types outside built-up area</td>
</tr>
<tr>
<td>- all roads (outside and inside built-up area)</td>
</tr>
<tr>
<td>4 ISA combination (additional functionality?)</td>
</tr>
<tr>
<td>- no ISA</td>
</tr>
<tr>
<td>- ISA warning</td>
</tr>
<tr>
<td>- ISA limiting</td>
</tr>
<tr>
<td>5 costs (purchase costs)</td>
</tr>
<tr>
<td>- Nlg.0 (for free)</td>
</tr>
<tr>
<td>- Nlg.1500</td>
</tr>
<tr>
<td>- Nlg.3000</td>
</tr>
<tr>
<td>6 insurance premium reduction</td>
</tr>
<tr>
<td>- 0%</td>
</tr>
<tr>
<td>- 25%</td>
</tr>
<tr>
<td>- 50%</td>
</tr>
<tr>
<td>overall utility (regression intercept)</td>
</tr>
</tbody>
</table>

Drivers’ Stated Preferences for the SAVE Systems
large part of the respondents did not realise that. The last three attributes provide insight in the possibility to increase the SAVE system's preference. The part-worth utilities estimated for the 'ISA combination' attribute indicate that an ISA with a 'warning only'-functionality is about equally well preferred to the absence of an ISA system, while an ISA that automatically limits the vehicle's speed to the maximum speed limit is strongly disliked. Furthermore, the SAVE's utility almost linearly increases with decreasing costs. Finally, utility increases with increasing assurance premium reductions. However, a reduction of 25% has practically no effect, while a reduction of 50% has a relatively large effect.

In this study the possible influence of the respondents background characteristics on their stated preferences has been researched. It appeared that the respondents' characteristics regarding sex, age, education and respectively annual mileage did not significantly influence their preference behaviour for SAVE (Van Hoytema, 1999). Hence, differences in preferences could not be explained by the individual background characteristics of individual respondents.

5.3 Attribute importance

The difference between the lowest and highest part-worth utility of the levels belonging to an attribute is often considered as an indication of attribute importance. This range may be expressed as the contribution to the sum of the ranges across all attributes, which is reported in the last column of Table 4. The figures indicate that 'purchase costs' is the most important attribute, closely followed by 'control'. This is at some distance followed by 'ISA-combination' and 'detection', while 'type of roads' and 'assurance premium reduction' are clearly the least important attributes. However, care must be taken in interpreting this indication of importance for at least two reasons. First, the magnitude is largely dependent on the levels chosen to represent the attribute. For example, if Nlg.2000 would have been chosen as the highest purchase cost level instead of Nlg.3000, 'control' probably would have been the most important attribute. Second, importance is based on the overall model, where different results may have been received if this was based on completely disaggregate models. The reason is that respondents may differ more on some attributes than on other attributes, and therefore their effects may be more levelled out if they are aggregated first.
5.4 Predictive validity of completely disaggregate models

The most severe test of a model's predictive validity is to test its ability to correctly predict the consumers' behaviour on real markets. As the SAVE system cannot be purchased yet, this test cannot be performed. Therefore, we test the models' ability to predict the observations for new choice alternatives, that is, the ratings observed for the two holdout profiles. To that effect, complete disaggregate preference models have been estimated in addition to the overall model just presented, that is, a separate model was estimated for each respondent. For each respondent we then predicted his holdout profile ratings based on his preference model. As an indication of the prediction accuracy, the absolute difference between the predicted rating and the observed rating for the holdout profile is calculated. The results averaged across the respondents are presented in Table 5. This table indicates that the mean absolute prediction error is about 1.2. This means that on average, the individual models make a prediction error of slightly more than one point on a ten point rating scale. As this is well above chance level, the predictive validity of the estimated preference model may be considered satisfactory.

Table 5 Mean absolute difference between observed and predicted holdout ratings

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holdout 1</td>
<td>1.21212</td>
<td>1.12447</td>
</tr>
<tr>
<td>Holdout 2</td>
<td>1.25469</td>
<td>0.96891</td>
</tr>
<tr>
<td>Total</td>
<td>1.23341</td>
<td>1.04636</td>
</tr>
</tbody>
</table>
In this paper, the stated preferences of drivers for the SAVE system have been examined by applying conjoint analysis. A utility function was estimated based on the respondents' ratings for 16 hypothetical profiles that each varied six functional and cost-related attributes. This utility function describes the part-worth utility contribution of each attribute level to the overall preferences of possible SAVE systems. From the estimated part-worth utilities, we can draw the following conclusions. The fact that, in this study, the utility increases with increasing number of causes that can be detected, implies that, contrary to expectation, drivers on average are favourable towards detection of alcohol and drugs abuse. This might be related to the fact that although drivers may not like that they are monitored once they have been drinking or taking drugs, the control attribute (to be discussed next) indicates that they appreciate control is taken over once they exceed a certain limit above which their driving behaviour increase the probability of accidents. The discussion, no doubt, will concentrate on the height of this limit. Moreover, people may also realise that the behaviour of other drivers is also detected which improves safety feelings. From the evaluation of the 'control' attribute can be concluded that drivers prefer to remain in control over their vehicle as long as possible, but they appreciate that the system takes over control once they are no longer able to maintain control themselves. This finding is consistent with the results of related studies that drivers, in general, are reluctant to technologies which actively intervene in their vehicle driving task (e.g. Hoedemaeker, 1999). From the evaluation of the type of roads attribute can be concluded that SAVE's utility increases if it functions on more road types. Moreover, drivers consider it especially important that SAVE functions on motorways. This makes sense, because driving at motorways is monotonous which may increase inattention. Moreover, as one drives at high speeds on motorways the consequences of a possible accident may be severe. The results for the 'ISA combination' attribute pointed out that the SAVE system preference cannot be increased by adding an Intelligent Speed Adapter for the same price. Purchase costs turned out to be the most important attribute, from which may be concluded that SAVE's preference can be heavily influenced by government subsidies. Finally, assurance premium reduction must be considerable (well above 25%) to produce any increase in the system's preference.
The explained variance of the overall utility model indicated that large differences exist between individual drivers. Although not reported, we examined whether these differences could be explained by differences in the drivers' personal characteristics, like sex, age, educational level, health effecting driving, car ownership, and yearly driven kilometres. This appeared not to be the case: The explained variance did increase significantly if the utility function was broken down by these characteristics. Apparently, other personal and driving related characteristics may explain the differences, like, for example, risk behaviour, driving style, perceived driving freedom, attitudes toward alcohol consumption and driving, traffic accident history, etc. Examination of these effects will be the subject of future research.

The part-worth utilities reported in this paper could be interpreted well, which gives face validity to the preference model. Furthermore, we concluded that the predictive validity, measured by the individual models' ability to predict the holdout profiles, was satisfactory. Hence, about these performance measures for conjoint modelling, the conjoint approach appears a useful tool to examine consumers' preferences for innovative transport technologies. In future research the usefulness of this approach will studied further on.
References


A Method to Reduce Excessive Noise Nuisance

TRAIL Research School, Delft, December 1999

Author
Ing. A.A.J. Naarding
Faculty of Technology Policy and Management, Delft University of Technology

Thesis supervisor
Prof. dr. R.E.G.M. van der Heijden
Faculty of Technology Policy and Management, Delft University of Technology

TRAIL Research School, December 1999


A Method to Reduce Excessive Noise Nuisance

TRAIL Research School, Delft, December 1999

Author
Ing. A.A.J. Nederveen
Faculty of Technology, Policy and Management, Delft University of Technology

Thesis supervisor
Prof. dr. ir. R.E.C.M. van der Heijden
Faculty of Technology, Policy and Management, Delft University of Technology
Abstract

In this paper a financial model will be presented to solve this problem. The basic concept of the model is: decreasing the highest acceptable level of noise exposure. This can be done by increasing the noise reduction rate at the local level and by investing in noise abatement measures.

1. Introduction .......................................................... 1
2. National policy .......................................................... 3
3. The scale of the noise issue ............................................. 9
4. A proposal for a new instrument to reduce excessive noise levels. 12
5. Case interim-period North Branch Betuweroute ...................... 15
6. Final remarks ................................................................ 17

References ..................................................................... 18
Abstract

After a decade of noise reduction policy, there are still situations which go beyond the maximally accepted noise level. Most of these situations are hard to solve. The standard solutions such as noise isolation have already been applied, but the noise reduction was too small. Money for expensive solutions is not available yet. Simply demolishing houses is unacceptable. From the resident's point of view it is the last alternative, from a financial point of view it can be very expensive and from the urban planning viewpoint losing housing might harm the city structure. Residents are lobbying and campaigning for years to find support and funds to solve their problems. It is uncertain if and when they will be successful.

In this paper a financial model will be presented to solve this problem. The basic concept of the model is: 'exceeding the highest acceptable level of noise costs money'. The owner of the infrastructure has to pay. He can decide to forward the bill to the infrastructure users. The collected money will be put into a special fund. After several years there will be enough money to finance the badly needed noise reduction works. The model will be discussed. An overview of advantages and disadvantages will be given. This model will not solve all noisy situations. In some occasions no realistic solution is possible. So saving money in that case does not make sense.

Due to growing traffic volumes and new traffic flows new noisy situations may be introduced. Well-known are the rapidly growing noise levels at the Dutch National Airport. Also the Dutch Rail-cargo line (Rotterdam-Germany) may cause a similar problem. The central line is under construction right now, but it is unclear when the supporting northern stretch will be realized. How the noise model will interfere in this discussion will be exposed.

Finally, the conclusion will be drawn if this model can improve the quality of life.
After a period of intense meditation and prayer, we find ourselves in a world that is truly beyond our comprehension. The experiences we have undergone are so profound that words fail to adequately describe them. Yet, we are driven to share these revelations with others, to help them understand the deeper truths that we have discovered.

It is our belief that the world we live in is much more complex than we realize. The forces that govern it are not simple or predictable. We must be prepared to face challenges and obstacles that we cannot even imagine. Our only hope is to trust in the guidance of our higher selves, to follow the path of love and compassion, and to embrace the unknown with open hearts.

We are called to be activists, to stand up for what is right, to fight against the forces of darkness. But we must do so with love and understanding, with a heart full of compassion for all beings. Only then can we hope to create a world that is truly peaceful and harmonious.

In this journey, we are not alone. We are connected to all of life, to the very fabric of the universe. We are part of a wider tapestry, and our actions have consequences that extend far beyond our own lifetime. We must be willing to make sacrifices, to put the needs of others before our own, to work towards a better future for all.

The road ahead is not easy, but we must be steadfast in our commitment to love and truth. We must be willing to face our fears and doubts, to overcome our darkness and embrace the light. Only then can we truly be said to be living a life of purpose and meaning.
1. Introduction

The concern of the national government is focused on the quality of life and the protection of the environment (article 21 of the Constitution). This article is the foundation of the environment-policy. The third National Environment Plan (NMP 3) exposes the policy (Department VROM, 1998). The NMP 3 introduces eight environment-themes. On each theme the trends on soil, air, and water quality are described. The general picture of the last decade shows a reduction in the (growth of) emissions. The quality of the environment has been improved but the results are to small to be visible in the natural environment. Especially the goals for emissions of CO\textsubscript{2} and NO\textsubscript{x} and the reduction of excessive noise are way out of reach. Due to the growing car mobility the noise level in urban areas still increases. In 1990 half of the population suffered serious noise, caused by traffic as well as other noise sources. Five years later this percentage was lowered to 40%. This reduction was achieved by technical measures like smooth pavements on freeways, noise barriers and facade insulation. The noise pollution is caused by traffic, neighbors, military aviation, civil aviation, industry and railroads. On some locations the noise is caused by several sources.

Traffic noise is ruled by the “Wet Geluidhinder” (Noise Act). In this act preference and maximum standards of noise levels on house facades are presented. If a house facade does not meet the preference standard a procedure can be started to get permission to use the maximum standard. If even the maximum standard on the facade can not be achieved, facade insulation must be applied to realize an acceptable noise level inside the house. The national government has made a facade insulation program to solve these cases. This program expires in 2008. In that year all houses should have an acceptable noise level inside the house. But by facade insulation the street noise level will not be reduced.

Noise pollution can be an issue at present situations and during the process of planning new infrastructure. New infrastructure (road or railroad) can introduce new locations that do not meet the noise standard. The “Wet Milieubeheer” (Environment Management Act) dictates that the impact on noise pollution must be part of the comparison of infrastructure alternatives. Most infrastructure works are motivated by solving an existing problem of capacity or environment. The Cargo line Rotterdam-Germany (Betuweroute) and the High Speed Rail are motivated by their impact on the modal split. The impact on this goal has an far higher weight during the process of selecting the best alternative then noise pollution. As a result new locations with an unacceptable noise level may be introduced and chances to solve existing locations can be spoiled. Consequently more emphasis on the impact of reduction of noise pollution is needed as mentioned in the NMP 3. This effort may not result in optimizing the impact on one issue (noise) because several other aspects contribute to the quality of life or the economic development of the country as well.
The structure of this article
This introduction has exposed that the policy of reducing excessive noise levels has not been successful. The goal of the NMP 3, no serious noise annoyed people in 2010, will not be achieved. The next paragraph (2) gives an historic overview of the national noise policy and an outlook of future developments. This overview is based on policy reports, national acts and scientific literature. The actual scope of the problem is described in paragraph 3. The number of houses that do not meet the standard is counted. An impression is given how much money the national government has spent and have to spent to reduce excessive noise levels. Based on the overview of paragraph 2 and the facts reported in paragraph 3 a new policy instrument will be developed. This instrument will be elaborated in paragraph 4. The new instrument will be applied on one case: the houses along the existing railroad line of the North branch of the Betuweroute (paragraph 5). An estimate will be made whether the new method will improve the local quality of life more effectively then the present policy. Some final remarks will be made about the usefulness of the new instrument (6).
2. National policy

The history of the national environment policy dates back to 1972. In that year the "urgentienota milieuhygiëne" was published (Department VROM, 1972). The report gives an overview of the impact on the environment of the growth of the population, of industry, consumption, urban sprawl and the mobility. The general environmental overview seems very bad. On a number of urgent issues a short term action program was proposed. The reduction of noise pollution was one of the issues. A special Noise Act was proposed. In this act noise standards must be set. Also an objective method was needed to calculate levels of noise. In solving the excessive noise levels, measures that reduced the emissions of noise were given priority. The new Noise Act was published in 1979 and came phasewise into force between 1979 and 1987.

Noise Act (1979)

The Noise Act is a very technical and detailed act. It gives noise standards as well as mathematical formulas to calculate noise levels. The new act presents two kinds of standards. The first and most restricted standard is the standard of preference, the second standard gives the maximum acceptable level of noise. The new standards had to be applied on every road with a daily car intensity of over 2450 cars. Below this figure, the number of cars is too low to generate unacceptable noise levels. Several years later this criterium has been replaced by the speed limit of 30 km/hour.

Standards have been set for noise levels inside the house and at the facade. Also a difference exists between daytime, evening and nighttime levels and between existing houses and houses to be constructed. Table 1 gives an overview of the noise standards. The two most important standards are the preference standard of 50 dB(A) and the maximum standard of 70 dB(a). The 50 dB(A)-standard is a level of noise that only few people considers as annoying. The ideal is that at no location, this standard will be exceeded. The 70 dB(A) standard is really the highest acceptable level. All locations with houses above this limit should be solved as soon as possible by measures like noise barriers or facade insulation.

<table>
<thead>
<tr>
<th>Road</th>
<th>House</th>
<th>urban street level</th>
<th>non urban street level</th>
<th>inside the house</th>
</tr>
</thead>
<tbody>
<tr>
<td>present**</td>
<td>present</td>
<td>70</td>
<td>70</td>
<td>45</td>
</tr>
<tr>
<td>planned</td>
<td>present</td>
<td>65*</td>
<td>65*</td>
<td>40*</td>
</tr>
<tr>
<td>present</td>
<td>planned</td>
<td>65</td>
<td>65</td>
<td>35</td>
</tr>
<tr>
<td>planned</td>
<td>planned</td>
<td>60</td>
<td>55</td>
<td>35</td>
</tr>
</tbody>
</table>

* These standards have been lowered in the late '90. The reduction was 0, 5 or 10 dB(A) depending on the road category and the location.

** Present stands for constructed before march 1986.
The same amount of noise is considered to be more annoying at nighttime as during daylight. This element has been taken care of, in calculating the noise levels. During nighttime 10 dB(A) must be added to the calculated noise emission. The new value must be used in Table 1. The nighttime period is defined as the period between 23:00 and 7:00 hours. Houses must obey the daytime and nighttime standards.

The Noise Act anticipates on transport innovations. New techniques can reduce the level of noise emissions. A reduction of 5 dB(A) is expected. This improvement has been absorbed in the standards. All standards are raised with 5 dB(A). In the years to come the bonus will be reduced to 2 dB(A) and after that abolished. In 1998 the bonus on “new road outside urban areas” has already have the 2 dB(A)-bonus.

The Noise Act is based on the calculation of traffic pollution. Calculations can be made for present and future situations. Measuring traffic noise situations could be done in present situations. The calculation procedure consists of four phases. In the first phase the noise emission of trucks and motors will be converted to an equivalent emission of cars. In the second phase the emission of the cars will be calculated. Table 2 gives an overview of noise emissions of cars by some common speed limits and car intensities. The table shows clearly that doubling the intensity result in an increase of the noise level by 3 dB(A).

<table>
<thead>
<tr>
<th>Intensity</th>
<th>50 km/hour</th>
<th>70 km/hour</th>
<th>100 km/hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>71.5</td>
<td>74</td>
<td>79</td>
</tr>
<tr>
<td>1000</td>
<td>74.5</td>
<td>77</td>
<td>82</td>
</tr>
<tr>
<td>1500</td>
<td>76.5</td>
<td>79</td>
<td>84</td>
</tr>
<tr>
<td>2000</td>
<td>77.5</td>
<td>80</td>
<td>85</td>
</tr>
</tbody>
</table>

In the third phase the additional emissions caused by tyre contact and reflection against opposite housing is summed up. All these effects together might result in an addition of 6 dB(A). In the final phase the reduction by dispersion and extinguishing is calculated. Table 3 shows the relation between distance and noise reduction.
Table 3  Maximum reduction of noise levels (in dB(A)) by distance

<table>
<thead>
<tr>
<th>distance</th>
<th>Reduction of noise (in dB(A))</th>
</tr>
</thead>
<tbody>
<tr>
<td>house</td>
<td></td>
</tr>
<tr>
<td>4 m.</td>
<td>9</td>
</tr>
<tr>
<td>10 m.</td>
<td>14</td>
</tr>
<tr>
<td>50 m.</td>
<td>25</td>
</tr>
<tr>
<td>100 m.</td>
<td>30</td>
</tr>
<tr>
<td>200 m.</td>
<td>34</td>
</tr>
</tbody>
</table>

Noise barriers have an impact comparable to the enlarging the distance between the road and the houses. Given this fact, barriers are more effective on houses close to the road. Right behind the barrier even an silent area can be made.

The standards and the calculation of noise caused by railroad traffic is different from the road formulas. The railroad calculation method is described in the Noise Act and the BGS (Besluit Geluidhinder Spoorwegen; Railroad Directive Noise Pollution). Surprisingly some railroad noise standard exceeds the road standards. The higher standards are motivated by German research which show a higher level of acceptance of railroad noise (Springer, 1991). This can be explained by psychological arguments (de Boer, 1994):

- railroad noise has a short time exposure with long intervals
- each train causes the same kind of noise
- the noise periods are scheduled (timetable)
- the train drivers cannot reduce or increase noise levels by driving behaviour, while car drivers can.

It can be expected that this advantage will disappear if the railroad tracks will be used more frequently. At high occupied tracks the noise emission profile becomes comparable to freeways.

Table 4  Summary of noise standard of railroads (in dB(A))

<table>
<thead>
<tr>
<th></th>
<th>preference standard street level</th>
<th>maximum standard street level</th>
<th>maximum standard inside the house</th>
</tr>
</thead>
<tbody>
<tr>
<td>until januari 2000</td>
<td>60</td>
<td>73</td>
<td>37</td>
</tr>
<tr>
<td>after januari 2000</td>
<td>57</td>
<td>70</td>
<td>37</td>
</tr>
</tbody>
</table>

Close to railroad stations higher limits are accepted.
Based on the railroad timetable of passenger trains, cargo trains and maintenance trains the noise emission of the railroad network is calculated. The results are written down in the “akoestisch spoorboekje” (acoustic timetable). The noise impact of changes in timetables, used train types and other railroad track constructions can easily be calculated.

**National Environment plan (1989)**
17 Years after the “Urgentienota milieuhygiëne” the first National Environment Plan is published (NMP) (VROM, 1989). The making of this plan was triggered by the report “Zorgen voor Morgen” (RIVM, 1988) which pointed out serious environmental problems. The new NMP was the first report with a comprehensive environment policy. Before 1988 environment policy was fragmented in several issue reports and parts of other policy reports.

**Second National Transportation Plan (1990)**
The “Tweede Structuurschema Verkeer en Vervoer” (SVV II, 1990) is the foundation of the national transport policy. The scope of the plan is the national transport infrastructure. The national goals do not affect local and regional policy. The central issue of the plan is the reduction and manipulation of mobility in stead of enlarging the freeway network. Points of concern are traffic safety, noise pollution, and carving up of the landscape. Most efforts are concentrated on reducing the growth of car use and stimulating alternative modes of transport, especially public transport. The report gives measurable goals which must be achieved in 2010. Two noise pollution goals are mentioned. Both goals compare the situation in 1987 with 2010.
- The area with a traffic noise level over 50 dB(A) may not increase
- The number of houses with a facade noise level of over 55 dB(A) must be reduced with 50%.

Emphasis is given to the reduction of noise emissions. Mentioned are noise restrictions to car types (European standards), electric engines, concentration of traffic on fewer roads, smooth pavement (ZOAB), speed reduction and speed control. A reduction of the speed limit on freeway from 120 to 100 km/hour results in 2 dB(A) emission reduction. Modern trains have a emission 5 to 10 dB(A) lower than the trains in service today. Reducing the volume of transport is highly futuristic. Even the SVV II is based on growing traffic volumes. It’s not clear how much money will be spent on excessive noise reduction.

**The Perspectives Report (1999)**
The first sketch of latest National Transportation Plan is presented in the “Perspectievennota” (Perspectives Report, 1999). The evaluation of the SWV-II showed that the targets mentions could not be achieved in 2010\(^1\). Mobility and emissions have increased more rapidly as expected. An change in modal split from car use to public transport is invisible. The report spends ten lines on noise caused by road and railroad traffic. The areas polluted with traffic noise over 50 dB(A) is stable but the number off annoyed people is still growing.

---

\(^1\) The Long term program Infrastructure and Transport 1998-2002 (MIT) seems unfounded optimistic without reason. The easy interim target (growth of pollution is allowed) is achieved, but the 2010 goal (serious reduction) will not be achieved.
Third National Environment plan (1998)
Also the third national environment policy plan concludes that the noise reduction target will not be achieved. This report takes all sources of noise into account. The report shows that the number of noise polluted people is decreasing slowly. The national target was no seriously noise annoyed people in 2010. In 1996 the percentage serious annoyed people is 30%. The report sums up the noise reductions proposal of the SVV II, and gives some additional measures: optimal design of infrastructure in urban areas, reduction of rush time peaks, telework and manipulation of car buying behavior.

<table>
<thead>
<tr>
<th>Source</th>
<th>annoyed</th>
<th>serious annoyed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road traffic</td>
<td>30%</td>
<td>25%</td>
</tr>
<tr>
<td>Railroad traffic</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>Military aviation</td>
<td>21%</td>
<td>10%</td>
</tr>
<tr>
<td>Civil aviation</td>
<td>5%</td>
<td>6%</td>
</tr>
<tr>
<td>Industry</td>
<td>5%</td>
<td>6%</td>
</tr>
<tr>
<td>Neighbors</td>
<td>34%</td>
<td>13%</td>
</tr>
</tbody>
</table>


Special attention is given to the report “modernization of the noise policy (MIG)”. In this report is proposed to abandon the Noise Act en modify the Act Environment Management.

Modernization Instruments Noise Policy, 1998 (MIG)
In 1985 the Noise Act has been evaluated. This has triggered a limited number of act modifications. In 1996 a group called “Marktwerking, Deregulering en Wetgevingskwaliteit (MDW: Marketing, Deregulation an Acts)” has again studied the Noise Act. The group concluded that this act had a positive contribution to the reduction of noise levels in urban areas. But the act has caused three unwanted side effects:
• Noise policy is judged as a stand alone policy. Discussions are based only on the noise reduction impact without judging the overall quality of life.
• The procedure to achieve the maximum standard is used much to often. No plans are based on the preference noise level.
• A stimulus to reduce the noise level is areas with an noise level between the preference and maximum level is missing.
Additionally, the group states noise to be a local problem which must be solved locally. The proposal in the MIG-report are closely related to the findings of the MDW-group. The most important elements are:

- The method to calculate noise levels will be kept out of the act. They will be maintained by an independent institute.
- The national government sets the preference and maximum standard. Local governments are free to set their own standards within this range.
- The local standard will be fixed in local plans. The people affected, have a right to participate in the setting of the local noise standards.
- The provincial authority checks the local noise standards in relation to the total local urban planning policy.
- The national and provincial government set their own standard for their infrastructure.
- Noise of railroads is limited to a maximum emission level.
- Locale governments make plans the reduced excessive levels of noise.

The formalization of this plan can only be made by a change of acts. This means than far after the year 2000 the new act will be operational.
3. The scale of the noise issue

The national government has spent 1.6 billion guilders to reduce rail and railroad noise pollution between 1979 and 1983 (Dietz, 1994). About 107,000 of the 400,000 houses with an noise level on the facade over 65 dB(A) have been solved. In 2/3 of the cases facade insulation was the most effective solution, the other cases were solved by noise barriers or constructing new roads. In 1993 about 300 km of Dutch freeways and 15 km railroad track was applied with noise barriers or noise walls.

In the period 1995-1998 the department of VROM has asked local governments to inventorise the number of houses that do not meet the noise standard. VROM has asked to describe the situation of 1987 (!) and to make a estimation for the 2010 (roads) and 2005 (railroad). The list has been divided in three categories:

- houses with a noise level over 65 dB(A) on the facade caused by road traffic. This list is called the A-list
- houses with a noise level between 60 and 65 dB(A) on the facade cause by road traffic. This list is called the B-list
- houses with a noise level about 65 dB(A) cause by railroad traffic. This list is called the rail-list

The national government has planned to insulate all houses on the A-list and Rail-list before 2010. The inventorization has resulted in a list of about 100,000 houses (see table 6). Houses that have been insulated between 1987 and 1998, should not be on the list. Also houses that could be insulated by Rail21-projects or project on the long term infrastructure planning list (MIT) should not be on the list.

| Table 6. Number of houses known by the Department of VROM on the A-list or the Rail-list with a noise level on the facade above 60 dB(A) of road noise or 65 dB(A) of railroad noise in 1987 |
|---------------------------------|------------|------------|------------|------------|
|                                 | Road 1987  | Road 2010  | Railroad 1987 | Railroad 2005 |
| 44 to 60 dB(A)                  | 3398       | 6202       | 0           | 1544       |
| 61 to 65 dB(A)                  | 35721      | 28705      | 9           | 3039       |
| 66 to 70 dB(A)                  | 39036      | 44252      | 15046       | 6084       |
| 71 to 75 dB(A)                  | 1004       | (*)        | 6664        | 7580       |
| 76 to 80 dB(A)                  | (*)        | (*)        | 1115        | 3892       |
| 80 to 91 dB(A)                  | (*)        | (*)        | 205         | 900        |
| Total                           | 79159      | 79159      | 23039       | 23039      |

(*) The inventory is done by local governments. They were not allowed to put houses on the list with facade noise levels above 75/70 dB(A). As a result these house are not on the list or are counted by a lower category.

Houses which will insulated by Rail21-projects or MIT-project are not on the list


TRAIL Research School, December 1999
Solving the A-list might cost 1 billion guilders; solving the Rail-list might take 250 million. The yearly budget is about 60 million. VROM does not expect all houses to be insulated before the target year 2010. The budget is too low and above all, the way of financing will be changed in 2003. The noise budget will be merged into urban revival fund. Local governments are free to spend the money on every project that improves the local quality of life. Houses on list B will never be insulated. A standard facade has an insulation of 22 dB(A). So the noise level inside a house will be close to the inside standard (see table 1).

The “Nationale Milieuverkenning” (National Environment Scan) 1997-2020 of the RIVM show different figures. The RIVM estimated that until 2020 about 125,000 houses with a facade level over 65 dB(A) will be insulated. It is unclear on which inventory this estimation is grounded and how many houses will be left over. The “Achtergronden bij de milieubalans” (background document; RIVM, 1996) gives an geographic overview (see picture 1). On this map all sources of noise have been taken into account. The picture gives an underestimation of the reality. Only the busiest 30,000 km of roads have been taken into account. So the noise pollution of many local main roads is missing. The noise level on the facades has been corrected to the subjective noise awareness. This new unit is called the Environment Quality Unit (MKM). Aspects that contribute to the MKM are the source of noise and the time of the day. The picture shows that the national airport (Amsterdam-Schiphol) and most of the military and regional airports exceed the 65 MKM. Also some freeways in de Randstad exceed this limit. At the 50 MKM level the complete national freeway network and most of the railways are visible.

If all the results of the inventories are compared, it is clear that no source gives an overview of the total number of house that exceeds the standards of 60 or 65 dB(A) in 1999. It must be at least 100,000 houses but probably much more. How many houses that have been insulated but have a facade noise level over 65 dB(A) are unknown. It is unlikely that all houses will be insulated before 2010.

The noise level on the facade can only be solved by source measures or noise barriers between the source and the houses. The RIVM expects a reduction of emissions by technical car and truck innovations. The emission reduction of new cars will be half a decibel and of new trucks 6 decibel. The smooth pavement (ZOAB) result in a reduction of 4 dB(A). All these reductions will be offset by the growth of car mobility.
All freeways in urban areas have already been applied by smooth ZOAB-pavement and noise barriers. Along railroads still many more noise barriers could be placed and noisy outdated trains could be taken out of service. It is very unlikely that with all these measure the 65 dB(A) limit will nowhere be exceeded. Not exceeding the 50 dB(A) sounds like utopia. Endlessly tolerate facades noise level above 65 dB(A) is not acceptable. On the streets this level is considered to be very annoying to many people and to residents the level is too high to sleep undisturbed with windows opened. Additional measures are needed to solve these problems.
4. A proposal for a new instrument to reduce excessive noise levels

By continuation of the present policy the national goal of no seriously noise disturbed people in 2010 will never be achieved. So more efforts or other policy instruments are needed. A widely accepted categorization of policy instruments is made by van der Doelen (1993). It can be an assistant to find effective instruments. Van der Doelen has made three categories:

- **juridical instruments**
  The Noise Act is an example of a classical juridical instrument. This act serves only one policy item (noise pollution reduction). The recent acts try to judge noise pollution in relation to other aspects that contribute to the quality of life. The new Environment Management Act is a good example.

- **economical instruments**
  Environment taxes and environment subsidizes are economical instruments to change behavior. According to Vermeulen (1994) environment-policy taxes can be very effective. But also less successful attempts have been made like an environmental motivated tax on car petrol and noisy airplanes.

- **communication instruments**
  Communication, explanation, discussion and treaties are four examples of this instrument. “Communication can be seen as the key to effective environment policy” (Winsemius, 1986). It contributes to the wishes and expectations of all involved actors. Unfortunately it takes very much time and the results are without engagement.

Today the development of the juridical instruments is in full swing. The tendency of the proposals is clear. The budget for noise reduction will be merged in the quality of life fund and noise impacts of measures will be compared to other aspects that contribute to the quality of life too. This might even result in a worsening of the noise levels but improving other aspects of the quality of life. The power of economic instrument will be reduced because the special noise fund will disappear and it is unlikely that the urban revival fund will give top priority to reduce noise pollution. Surprisingly no innovations are expected on the communication level.

The new instrument that will be elaborated in this article is an economic instrument. This instrument can be effective if it meets a number of criteria. Vermeulen (1994) mentions the next constraint:

- **Conditionally** The charge expires if the noise pollution is solved.
- **Proportionally** The charge level depends on the level of exceeding the standard.
- **Substantially** The charge is high enough to make alternatives the cheaper choice.
- **Recognizable** The charge known to the people.

To make the instrument feasible two additional constraints have to be met:

- **Measurable** The annoyance must be measured objectively.
- **Elaboration costs** The higher the number and variety of sources the higher the elaboration costs.
In the Noise Act (1979) noise charges were introduced. One charge was linked to petrol prices and car ownership. The charge was operational between 1981 and 1987. Yearly it generated about 75 million up to 90 million guilders. In 1988 the charge was replaced by a general petrol charge. The money of this charge was added to the national budget. Both charges had a very small impact on reducing noise pollution. It did not meet any of the constraints of Vermeulen.

The new instrument has been constructed according to the constraints. The new instrument targets on reducing noise pollution levels on facades over 65 dB(A) caused by road or railroad traffic. The sources of noise will be charged. Each car and train at a specific location or, much easier to implement, the organization that can represent them, like road and railroad track owners, will be charged. The collected money will be put into a location related fund. After several years enough money has been gathered to finance noise reduction measures. As soon as the measures have been implemented and the noise excessive pollution has been solved the change on that location ends. The charge will be in favor of every house that exceeds the maximum standard. The level of charge will go up with the number of decibels that exceed the noise standard. This can be done by calculating the noise levels and checking the results by field measures.

The charge level is very important. If the charge is low, it will take decades before enough money is collected to finances solutions. This conflicts with the goal to reduces excessive noise pollution within a reasonable time. An extreme high level of charges is unacceptable from the viewpoint of car and train users. On the extreme side a charge that high, means that the infrastructure is to expensive to use. It leads to changes in route choice, location choice of houses and companies and modal split.

The charge level can be calculated from two viewpoints:

- The existing budget of reducing noise pollution of about DFL 60 million. The yearly sum of all charges will be equal to this budget. This way of financing is easy to implement because it is close to the present policy. The impact of this way of financing is small. It does not generate additional funds. The only result is a redistribution of a budget that is known to be to small to reduce excessive noise.
- The cost of needed measures. Noise pollution can totally be solved by taking away the source or the houses. In practice this leads to constructing tunnels, modifying traffic flows to other (new) roads or railroads or demolishing houses. Out of these alternatives demolishing houses is probably the cheapest with the smallest range in costs.
Based on the average house price of about DFL 300,000 and a fund generation time of 25 years, each house with 1 dB(A) facade noise level above the standard will generate a yearly charge of 12,000.-. This charge must be divided by all passing cars or trains.

The new economic instrument has a number of advantages above the common practice:

- All existing locations with an excessive noise level will be solved sometime. Sooner or later enough money will be collected in the fund to finance an acceptable solution. On locations with a high level of noise pollution, the fund will be filled more rapidly than at other locations.
- The resident’s policy to demand the most expensive solution is less interesting. It leads to more years waiting time before enough money is collected. Cheaper but effective solutions become more attractive.
- The impact on noise pollution on existing locations becomes more interesting when new road or railroad infrastructure is planned. New infrastructure which contributes to reduce excessive noise levels elsewhere becomes more attractive. Creating new locations with excessive noise level will immediately be charged.

Also disadvantages are already visible:

- In the beginning the National Government pays all the charges. All the roads and railroads are state property. The Government has to find a way to charge the users that cause the excessive noise. So an easy system of levy toll is needed. A general petrol tax is unwanted because the relation between petrol and excessive noise on a specific location is too weak.
- As long as the fund is too small to finance a solution nothing will happen. All funds have the same starting year and need several years to fill up. So no location will be solved the years.
- It’s is very difficult to change the 65 dB(A) standard to a stricter value like 60 dB(A). Locations which already has been reconstructed will probably not meet the new standard.
- It takes years to generate enough money for expensive solutions. This will stimulate the policy to accept cheap solutions that do solve the problems partly and ruin the visual quality.
- What must be done on locations that cannot be solved? Demolishing housing is unwanted (all houses have an historical value, city structure will be ruined and the first row is the noise barrier for the houses behind). Noise barriers are unacceptable: It cause visual pollution and make the street opposite out of reach. Also the tunnel will not be possible. All the traffic has a destination on surface level, or the underground space it occupied with other infrastructure.
- Finally, very interesting is the debate that starts when enough money is generated for first cheap solution. Who decided if or which measure should to be taken? Is it the government, the charge payers or the residents. A procedure is needed to regulate possible conflicts.

---

2 Example: The charge on a location with 100 houses close to a busy road or railtrack costs 1,6 cent per car and DFL 4,56 per train from each decibel above the maximum standard.
In June 1995 to Dutch Parliament has decided to construct a dedicated cargo railroad line between Rotterdam and the German Border (Arnhem/Elst). This line is called the Betuweroute. The destination can be north Netherlands, north Germany or Scandinavia. Some people argue that an addition cargo line is needed between Arnhem and the north, to accommodate these cargo trains. The new line is called the North Branch of the Betuweroute. The National Government will decide in the end of 1999 which trace will be selected. After this decision has been taken, money to finance the project has to be found before construction can take place. The Betuweline is scheduled to open in 2005, the North Branch will be operational several yearly later.

According to the Raillist of VROM 1669 houses along the existing track Velp-Deventer-Oldenzaal, have an noise level on the facade over 65 dB(A) in 1987. About 40% is the group has a noise pollution above 70 dB(A) (in 2005 50%). This noise level is caused by passenger trains, a few cargo trains, and for 284 houses road traffic. After the year 2005 the first Betuweline cargo train will pass. Railned has estimated that 70 trains will pass daily in 2005, 91 in 2010 and 114 in 2015 (Department of public works, 1998). The number of houses between Arnhem and Deventer with a level over 65 dB(A) is counted. Too many houses receive unacceptable noise levels. In the report is proposed to construct 34 kilometer noise barriers between Arnhem/Elst and Deventer. The average height of the barriers is 2.1 meter. After this measure the number of houses that exceed the 65 dB(A)-standard is still 1900. Above that the noise barriers do not fit in the urban structure of Oosterbeek, Arnhem, de Steeg, Ellecom, Leuvenheim, de Hoven and Eefde. Only one village along this track is missing in this list.

Under the present act residents do not have any power to stop this process. Their houses exceed the maximum standard and in the future it will be worse, even after noise barriers have been applied. The government decision of the trace and environment report of the North Branch is very uncertain.

---

3 Very recently both governments are negotiating to change the Treaty.

4 Houses constructed after 1987 of insulated before 1998 have not been counted.

5 The track Elst-Velp is 17 km. Velp-Deventer is 39 km; Most noise barriers are planned between Velp and Deventer.
It's very realistic to expect that the existing track will be chosen, because the number of cargo train might be too low to motivate a dedicated cargo line. It is very likely the noise pollution will be resistant even after 2015. The residents can do nothing else then waiting until they are on the top of the list of the insulation program, praying the noise reduction budget still exists.

The new policy instrument gives an end time of the excessive noise. Each passing passenger train and cargo train must pay because together they exceed the maximum noise standard. The locations funds will be filled rapidly because the number of house is high, at least 1669 houses (Raillist, VROM), and level of exceeding the standard too. The highest level without the new cargo trains is already 84 dB(A). This charge triggers three impacts:

- Money will be generated to solve local noise pollution.
- It becomes less attractive to endlessly exceed noise standards. It is an stimulus to invest in silent locomotives, cars and track constructions.
- The impact of noise pollution becomes more important in the North Branch trace choice. Noise will also be an issue at the cost comparison.
6. Final remarks

Twenty years after the Noise Act has been published a large number of houses still do not meet the maximum standard. This poor result can be explained by the limited insulation budget, the rapid growth of the traffic volumes and a change in policy priorities inside and outside the environmental policy. It's is unlikely that the goal set for 2010 will be achieved. To reduce the number of seriously noise annoyed people an additional effort is needed. In this article is proposed to charge the sources of noise (cars and trains). The charges will fill up location related charge funds. As soon as a fund has enough money to finance a solution an very interesting discussion will start. Actors involved in this discussion are the road or railroad owner, local residents, and the local government. In contradiction to the present policy all actors share the same target: solving excessive noise levels by accepted solutions. By using this method all locations with an excessive noise level can be solved within 25 years.
References


- Boer, E. de, J.G. the Man, L. Molenkamp, Overweg met spoorwegen. Een handleiding met achtergrondinformatie en beoordelingsmethoden voor bewoners die hinder ondervinden of planning, aanleg en gebruik van spoorwegen in the woonomgeving (Solving problems caused by railroads. A manual with background information and methods to analyze problems caused by the planning, construction and use of railroads in urban areas). TU Delft, department of civil engineering, November 1994.


- Ministerie van Verkeer en Waterstaat (Department of Transportation), Tweede Structuurschema Verkeer en Vervoer, deel d, regeringsbeslissing (Second National Transportation Plan). Den Haag, 1990.

- Ministerie van Verkeer en Waterstaat (Department of Transportation), Perspectievennota Verkeer and Vervoer (The Perpectives Report), februari 1999.


• RIVM (National health and environment institute), Achtergronden bij de milieubalans 1996 (Background information about the environment 1996), Bilthoven, 1996.


• Winsemius, P., Gast in eigen huis; beschouwingen over milieumanagement (Guest in your own house; considerations about managing the environment). Alphen aan den Rijn, 1986.


A Method to Reduce Excessive Noise Nuisance
Introducing life cycle cost management in the railways: opportunities and drawbacks

TRAIL Research School, Delft, September 1, 1999

Author
Ir. A. Zoeteman
Faculty of Technology, Policy and Management
Delft University of Technology

Thesis supervisors
Prof.dr.ir. R.E.C.M. van der Heijden
Faculty of Technology, Policy and Management
Delft University of Technology
Prof.dr.ir. C. Esveld
Faculty of Civil Engineering and Geosciences
Delft University of Technology
Introducing life cycle cost management to the airline industry

Airline Research Series, Dell'Orto Press, 1990

Author

J. A. Storms

1. Cost of Life-Cycle, Fuel and Maintenance
   Cost of Use and Use

Title Page

Prepared by J. A. Storms for the U.S. Department of Transportation

© 1990 J. A. Storms

Dell'Orto Press 1990

Page 15
Abstract

Contents

1 INTRODUCTION

2 LIFE CYCLE COST MANAGEMENT

3 THREE CASES ON DEVELOPING DECISION-SUPPORT

3.1 Life cycle cost analysis at the Madrid Metro

3.1.1 Problem setting

3.1.2 Providing decision-support

3.2 Appraisal of the Embedded Rail Structure

3.2.1 Problem setting

3.2.2 Providing decision-support

3.3 Track structure selection for the HSL South

3.3.1 Problem setting

4 CONCLUSION: OPPORTUNITIES AND DRAWBACKS

REFERENCES
Abstract

In the last decade changing policy on railways has created a separate entity responsible for infrastructure management in many European countries. The consequence has been the requirement to shift to a performance-based approach for railway infrastructure management. The infrastructure manager has to cope with increasing demands of the central government and the transport operators. The functionality of the railway system must facilitate a more diverse traffic demand, while costs have to be reduced and availability has to be improved without endangering the railway safety.

Since the railway assets have rather long life spans, decision makers need to consider the long-term effects of decisions according to the 'life cycle management' concept. However, many obstacles occur in current railway practices, e.g. lacking data, distribution of risks, political and organisational factors. All these factors create an environment that does not stimulate life cycle cost based management.

In the development of decision support systems this environment should be taken into account. The experiences from three projects are discussed, which lead to the conclusion that the use of decision support systems could be promising in realising decision making that is more directed towards life cycle costs. The major contribution of a DSS can be expected by facilitating the discussion, not by prescribing a 'best solution'.
Aperçu

...
1 Introduction

During the 1990s the position and structure of the railway organisation has changed drastically in many countries. Policy makers increasingly consider railways to be an important part of the overall transportation system. In the European Union this change is illustrated well by the official EU transportation policy: railways should play an important role within the congested transport system. However, in order to realise that objective, the performance has to be improved on aspects such as availability, reliability, cost-effectiveness, safety and comfort.

In many countries, such as the Netherlands, policy makers have chosen to separate the railway operations from the infrastructure management. This 'vertical separation' should facilitate competition in the rail transport market and should distinguish responsibilities more clearly [EU 1991; Vincent et al 1996].

In the Netherlands the Dutch government has established three so-called task organisations: an infrastructure manager, NS Railinfrabeheer, a capacity manager, Railned, and an organisation for traffic control, NS Verkeersleiding. The outplacement of these organisations out of the organisation of Netherlands Railways (NS) is scheduled for the year 2000. For the management of the High Speed Line South, which is to be taken into service in 2005, yet another arrangement has been made: a private sector based Infrastructure Provider (IP) will be selected to construct and manage the railway system for a period of 25 years. The IP will receive a quarterly fee for the realised system availability – a penalty system will be applied for availability performance [HSL South ITC 1999].

The environment created by changing railway policy has an impact on the role of 'infrastructure management'. The infrastructure manager, responsible for design, construction, maintenance, renewal and upgrading of the infrastructure, has a clearly defined role and is confronted by increasing performance requirements of the other actors. Budgets are reduced, as availability has to be increased without endangering the traffic safety. A systematic approach is needed for communication with the capacity manager and central government and for guaranteeing defined levels of performance. This approach is lacking in an organisation where maintenance and renewal has long been planned and executed according to individual experience and skills [Swier 1997].

In order to establish a continuously high performance level of the infrastructure it is important to systematically consider the long-term effects of decisions: for instance, expected costs for different levels of performance (availability, etc.) must be made explicit for making agreements with the central government and the transport operators. Performance and costs must be linked in a transparent way in order to plan the efforts of maintaining the railway assets in relation to the performance required.

In this paper an approach that aims at supporting the decision-making process in order to realise an improved performance with lower, better manageable costs is proposed. The approach is based on the principles of life cycle cost management and decision-support systems (paragraph 2). In paragraph 3 the approach is illustrated in three cases, at the Madrid Metro, the Netherlands Railways and the Dutch High Speed Line South. Based on these cases the most important opportunities and drawbacks for life cycle cost management are identified in paragraph 4.
2 Life cycle cost management

Since railway assets have rather long life spans, are costly and have an important impact on maintenance and reliability, long term planning is a key in realising the objectives of availability, low cost infrastructure and reliability. Although daily decision making is very important for an adequate performance of the railway system, the overall cost and availability characteristics are closely linked to the decisions in design, construction and long-term planning of maintenance and renewals. The choice of a track structure or a power supply system determines the conditions under which train operations have to be performed for a long period, e.g. maintenance needs and number of failures.

This observation can be translated into a need to analyse the long-term performance of a system, in specific the so-called 'life cycle costs', as good as possible. The life cycle costs need to be optimised under a number of constraints, such as maintaining an adequate level of safety and comfort and environmental constraints. The life cycle costs are made up by costs of construction, maintenance, renewal (cost of ownership) and costs related to availability and reliability (cost of operation). Especially, the costs caused by disruption to traffic are becoming more important in situations, where railway services have to compete with other traffic modes e.g. high-speed and commuter lines.

Life cycle cost management is a concept that is aimed towards decision making based on these long-term effects. The life cycle management (LCM) concept is closely connected to the field of terotechnology [Van Duijvenvoorden and Verdoes, 1993]. Terotechnology has been defined as a combination of management, financial, and technical methods, applied to capital goods in order to reduce the life cycle costs.

In the LCM concept information on life cycle costs is the starting point for the decision-making process on different alternative solutions. The methods to collect and present this information are referred to as life cycle costing:

Life cycle costing is an economic assessment of an item, system, or facility and competing design alternatives considering all significant costs over the economic life, expressed in terms of equivalent currency units [modified, after Stephen and Dell'Isola, 1995].

Problems within the LCM concept are caused by the uncertainty of expectations on future infrastructure and operations performance, e.g. traffic intensities, maintenance and availability, especially in the assessment of innovative technology. For this reason in this paper an approach based on the ideas on 'decision support systems' (DSS) is elaborated. Decision-support systems are computer-based systems that produce information for supporting decision makers in making choices. The advantage of these systems is that different scenarios can be analysed in an easy way (what-if analysis) and the important factors in the evaluation of different alternative solutions can be identified i.e. the factors that influence costs or availability heavily (sensitivity analysis). In this way decision-support systems do not prescribe the best solution, but provide an indication of the range of possible outcomes and provide insight in the robustness of the alternatives. Moreover, the effects of new information can be tested during the decision-making process.

1 Comfort for passengers is e.g. partly determined by the geometric quality of the railway track.
3 Three cases on developing decision-support

As has been described in paragraph 2, it is tried to support the LCM concept by developing decision support systems that are able to analyse different scenarios for future use and maintenance on a railway line in order to support the choice between different alternatives for the railway system. The results of three cases are described; all cases concern the choice of a track structure, which has to be made in the process of designing a new railway line. Once operations start on the new railway line, changing the track structure is hardly possible without huge costs or disruption.

3.1 Life cycle cost analysis at the Madrid Metro

3.1.1 Problem setting

In Madrid a large extension program of the metro network, Ampliación 94/99, has been started in order to cope with the rapid increasing transport demand. The Directorate General of Infrastructure of the Madrid Regional Government (Comunidad de Madrid) is responsible for the construction of the new metro infrastructure. It concerns 35 kilometres of new lines. This is done in close co-operation with the Madrid Metro, which will take care of the exploitation and maintenance.

For the extensions so-called block track was initially selected by the decision-makers. Both actors doubted, however, whether this was the best structure for the Metro, considering the total (financial) performance. These doubts were caused by constructional problems resulting in unexpected costs. The question raised if block track was worth the higher investment.

In order to investigate the effects of using block track instead of the conventionally applied ballasted track a decision-support system was developed that was able to calculate the total expenses on the track [Zoeteman, 1998a]. Disruption of passenger services was not an issue, since the nights are available for maintenance.

![Figure 1: Ballasted track (left) and block track (right)](image)

3.1.2 Providing decision-support

The infrastructure costs were in the Madrid Metro situation the most important effect to be considered. For this purpose a model was developed for forecasting the life cycle costs on the planned lines for different scenarios.
From the start both Madrid Government, which asked for the analysis, and Madrid Metro were involved. With both actors the decisive maintenance activities were selected; the model was developed in a way that it could cope with the data available. For instance, technical maintenance standards were undefined - for this reason the data of the Annual Reports of the maintenance department of the Metro was used. It was possible to calculate productivity figures from these reports. Interviews with Metro staff were used for the collection of data on labour costs, which proved to be a decisive element in the total costs. In fact, all maintenance and renewal had to be performed in the nights. This makes mechanised renewals hardly possible and causes inefficiency (mostly due to short working nights).

In a next step the Government and the contractors provided data on the construction costs. Finally, a maintenance plan for the new lines was developed in co-operation with the maintenance staff of the Madrid Metro.

Although a large amount of uncertainty was caused by shortcomings in the data management of the maintenance department, in all investigated scenarios block track proved to be at least 10% less expensive over 40 years, including the financing costs of the extra investment required. This was a clear answer to the question of the Madrid Government. The Madrid Government has decided to continue using block track, as recommended by TU Delft [Esveld et al., 1998].

The analysis had some unexpected side effects: the analysis showed the lack of accurate, accessible data at the maintenance department and the inconsistency in the data collected. This has caused disturbance in the organisation of Madrid Metro, which was clearly not foreseen at the start.

3.2 Appraisal of the Embedded Rail Structure

3.2.1 Problem setting

Within the European railways a discussion on ballasted and slab track systems has started, since many new (concrete and asphalt) slab track systems have been developed that require less maintenance. In the Netherlands the embedded rail structure (ERS) is considered a promising option. Its excellent maintenance performance has been proven in level crossings and a test track for over 20 years. The structure is unique for its continuously (elastically) supported rail - the only structure in the railways with this feature. This reduces the rail fatigue (caused by the loads of passing trains) and the rail is fixed for 40 or 50 years (on the national rail network) with little maintenance, according to lab experiments, simulations and field tests. Recently, a mechanised construction process with a so-called paver, has led to a drop in the construction cost, making the structure an attractive alternative for ballasted track.

Due to this new technique Strukton Railinfra, a large railway contractor, needed an investigation of the financial performance of the embedded rail structure (ERS) in order to know if ERS could be an attractive option for the railways. Also the effects on the train operations should be considered in decision-making on this issue.

3.2.2 Providing decision-support

Although the use of this structure (ERS) has an impact on many facets of the infrastructure, such as noise, construction methods, work force required, substructure, and risks, it was decided to concentrate in a first analysis on the costs of ownership and the traffic disruption level, leading to extra operation costs and income loss. The financial performance was considered to be a first factor in the appraisal of the new structure. ERS was compared with the conventionally applied ballasted track.
In co-operation with the technical experts and decision-makers at Strukton Railinfra a model was developed (figure 3). A different elaboration was chosen than in the Madrid Metro case [Zoeteman, 1998b]:

- A comprehensive database on the maintenance performance was not available. Since it concerned a relative new structure, only technical estimates were available.
- Maintenance and renewal at Netherlands Railways is to a large degree scheduled in large-scale operations; maintenance standards were available, such as the tonnage limit for the track components.

The data was collected for both traditional ballasted track and the ERS. For the ERS the costs of construction were extracted from the budget of a pilot project near the city of Best, the Netherlands. In 1998 embedded rail was constructed with a paver over a length of 3 kilometers, as a test for the coming High Speed Line South. Moreover, maintenance estimates of a research project of RIB were used. For the calculation of disruption a number of use scenarios were investigated, based on the international UIC classification of railway lines.

![Figure 2 Construction of Embedded Rail with paver](image)

<table>
<thead>
<tr>
<th>Table 1: Tonnage Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mainte nance slot length</strong>, work distribution, productivity rates</td>
</tr>
<tr>
<td>Unit costs of major overhaul and renewals, discount rate</td>
</tr>
<tr>
<td>Construction costs</td>
</tr>
<tr>
<td>Quality indicators, thresholds</td>
</tr>
<tr>
<td>Routes information</td>
</tr>
</tbody>
</table>

Disruption to traffic >> disruption costs

![Figure 3 Concept of the Life Cycle Cost Model](image)
In all scenarios the ERS proved to result in on average a 20% reduction of the life cycle cost over 45 year, a rather stable outcome: the higher investment pays out in lower total expenses on the track. Financing costs were included in this calculation using a discounting rate. Moreover, the disruption to train traffic proved to be negligible, which is a very promising outcome for demanding train services. Strukton Railinfra considers this result important and will develop its expertise on the structure further.

3.3 Track structure selection for the HSL South

3.3.1 Problem setting

The Dutch State has decided to tender construction and management of the railway system of the Dutch High Speed Line South (from Amsterdam to the Belgian border) to a so-called Infrastructure Provider (IP) for 30 years, which has to finance construction and maintenance in advance. The IP will get a quarterly fee, if performance on availability is adequate. A performance payment regime includes penalties, if system availability is below 99% per 28 days. The track structure is the main variable in the design process, which is yet under discussion. Since the track structure has to be financed by the IP, costs and risks are keywords in this discussion - especially, since it has to be financed by the private sector, the risks play an important role. A number of track structures are under consideration, such as the ballasted track, the ERS and a German and Japanese track structure. The Japanese structure, named Shinkansen slab track, has the least risk (proven technology) and requires few maintenance. It is however more expensive than the others.

The traffic frequency scheduled over the HSL South, the performance payment regime and the few slot hours available for maintenance cause a situation that is not comparable to the circumstances on the national rail network. It could be possible that only 4 hours are available for maintenance; this could be problematic for e.g. ERS, the renewal of rails will cause many maintainability problems causing high costs (it takes more time than on other structures). A final risk is that the substructure, below the track structure, is constructed independently from the track (by another consortium), but settlements will be part of the risk for the IP. This is especially problematic for the slab track structures, such as the German track and the ERS. A solution is to offer ERS with prefab slabs instead of with a paver (continuous slab; see figure 2), but this raises the construction costs again to a high level.

---

*Figure 4 Outcomes for a representative NS situation (present values)*

**Discounted life cycle cost expectation**

<table>
<thead>
<tr>
<th>Construction</th>
<th>Annual maintenance</th>
<th>Major overhaul and renewal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ballasted轨</td>
<td>ERS</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>400</td>
<td>600</td>
<td>800</td>
</tr>
<tr>
<td>800</td>
<td>1,000</td>
<td>1,200</td>
</tr>
</tbody>
</table>

---

2 This case is yet under way, but some findings can already be presented.

---

6 Introducing life cycle cost management in the railways
3.3.2 Developing decision-support

For the Alstom Strukton Volker TUC Rail ABN Amro HSBC consortium a decision support system has been developed that is able to calculate the total costs up to 2034 [Zoeteman, 1999]. Factors relating to the HSL South situation, such as penalty costs, scheduled trains and construction time, can be varied in the system - the performance of the track structures in many future situations can be simulated. The DSS is an extension of the earlier DSS (paragraph 3.2). The analysis in the HSL South situation can however lead to different outcomes, as described in paragraph 3.3.1. At the moment the DSS is used for what-if analysis in order to present to the State the range of outcomes and the effects of the proposed risk distribution, the maintenance regime and the performance payment. In the figures below some output is given.

Revenue distribution including interest (MIO EURO)

Figure 5: Example of output

Although this output, the annual fee needed by the IP for the different track structures, is just an example, it is interesting to see that the distribution of costs over construction, maintenance, penalties and risks differs a lot for the different track structures. Such a cost distribution is information that is very important within the consortium (for discussion between constructors and banks) and for the discussion between the Dutch State and the consortium. The consortium could for instance present the effects of a different performance payment regime or a different risk covering regime. The DSS proofs especially for this reason to be an important tool.
4 Conclusion: opportunities and drawbacks

The described cases have shown a way to support the concept of life cycle (cost) management. The cases showed a number of problems related to the LCM concept and a way to deal with these problems. A number of problems proof to occur in respect to the LCM concept:

Lacking, inaccurate or unreliable maintenance data
In the Madrid Metro case this was for instance shown by the problems related to collecting the maintenance data. Although both track structures analysed were in use for many years, a database on the maintenance performance was not available and the collection of maintenance data in general was very difficult. Maintenance has proven to be a neglected management area for a long period, not only at the Madrid Metro. A problem with the assessment of ERS and the German track structure (Rheda) proved to be the fact that both are innovative track structures. Estimates are available thanks to pilot projects, simulations and laboratory experiments, but the margins for the maintenance expectation are often larger than for the conventional, well-known structures.

Distribution of responsibilities and risks
The distribution of responsibilities and risks over different organisations cause problems in realising the LCM concept. Organisations that perform the maintenance often do not have an incentive to reduce maintenance, since it can be part of their job. The data needed to produce the information on life cycle costs is also distributed over these organisations. An opportunity in this respect is the IP contract, in which the private sector based manager is stimulated to optimise both construction and maintenance, in order to win the tender and to deal with the ‘performance payment regime’. An obstacle in the case of private sector involvement can be the risk covering; the private sector proves to be reluctant in the use of innovative technology.

Organisational factors
Often experiences and opinions within an organisation determine choices, e.g. some railways proof to stick to their track structure since they have developed it. In this way other track structures are not considered, although a life cycle cost analysis could lead to new information for the decision makers.

The decision-support systems described are a tool for the decision makers to get more insight and “feeling” for the long-term consequences of choices for the infrastructure performance. A DSS can certainly not take a way the uncertainty around these decisions, but can make it manageable. In this way it can support the decision makers and be a vehicle for the communication between different actors: the discussion amongst decision makers can be structured by the DSS; the DSS can for instance show the expected effects of a different risk regime and different availability requirements. This is needed in the situation, where a separate entity performs the infrastructure management and has to settle performance requirements in co-operation with the central government and the transport operators.
References


Duijvenvoorden, C. van and B.J. Verdoes, Principles van onderhoudsmanagement, Leiderdorp, The Netherlands, 1993 (In Dutch)


Stephen, K.J. and A. Dell'Isola, Life cycle costing for design professionals, McGraw-Hill, USA, 1995


Zoeteman, A., A financial comparison of track structures: application of life cycle costing for metro track construction in Madrid, Comunidad de Madrid, Spain, 1998a

Zoeteman, A., Onderzoek naar life cycle costs bij de Nederlandse Spoorwegen: de eerste resultaten van toepassing van LIFE CYCLE COST PLAN 98 op ballastspoor en de ingegoten spoorstaaf constructie, Strukton Railinfra, Maarsen, The Netherlands, 1998b (In Dutch)

Conclusion: opportunities and drawbacks

The developed case has shown how M
ds have improved the competitiveness of t
electronic commerce and the need for a
more efficient infrastructure. This has been
demonstrated by the case of M
ds, where t
had to face the challenge of providing a
for the electronic commerce market. The
of M
ds has been able to achieve a
increase in sales and profits, and to
the competition from other firms. Also,
the case of M
ds has shown how t
can use technology to create a
competitive advantage.

Organizations' factors

Organizations that are able to adapt to
the changes in their environment and to
the demands of the market are the ones
that are more likely to succeed. Organizations
that are not able to adapt to these
changes are at risk of becoming obsolete.

The decision support systems described
are a tool for the decision makers to get
more insights and "feel" for the long-term
capabilities of solutions for the
performance of a system. A DSM can
nearly tell us why the performance of
decision-making systems is at its best,
and it can also be a managerial tool. In this
way it can support the decision
makers and be a vehicle for the
communication between different actors. The
DSM can be structured in a way that it
has two basic components: a decision
making and a management. This
structured can be used to identify
the interaction of different actors
and to assess the performance of the
decision-making systems.
Integrated land-use & transport modeling

The structure of the land-use model

A growing population and intensifying social-economic activities in most fertile delta regions lead to a variety of resources (e.g. water, land) and other commonly known urban problems as air pollution or traffic congestion. For a long-term regional plan (including the impacts of transport measures on the spatial development) the spatial planning and the spatially related transport sectors have to be evaluated in an integrated way. Despite the need for such an integrated model and the research efforts of the last thirty years in the field of integrated land-use and transport modeling, the number of integrated models in practice is still very limited. Another important conclusion is that the principal model concepts of the lack of a separation process between land-use & transport modeling.

Author
Ir. B. Zondag
Faculty of Civil Engineering and Geosciences, Delft University of Technology

Thesis supervisor
Prof. Ir. F.M. Sanders
Faculty of Civil Engineering and Geosciences, Delft University of Technology
Introduction land-use & transport modelling

The structure of the land-use model

TRAIL Research School Final December 1999

Author
IL. S. Campbell

Scality of CHI and Development of Operational Case Studies of Transportation

Transportation
Planning & Strategy

Faculty of Urban and Regional Planning, University of Waterloo
Abstract

This paper discusses the research objectives of my dissertation research project. The dissertation study is in a preliminary stage and has started in March 1999 and the plan is to finalize the study in 2003. The dissertation study continues a line of research in the area of integrated land-use and transport modeling.

A growing population and intensifying social-economic activities in most fertile delta regions result in a scarcity on resources (e.g. water, land) and other commonly known urban problems as air pollution or traffic congestion. For a long-term regional plan (including the impacts of transport measures on the spatial development) the spatial planning and the spatially related transport sectors have to be evaluated in an integrated way. Despite the need for such an integrated model and the research efforts of the last thirty years in the field of integrated land-use and transport modeling the number of integrated models in practical-use is very limited. Another important weakness beside the practical limitations is the lack of a calibration concept for complex land-use & transport models.

Based on comparative studies and own ideas a spatial model structure will be developed. The main challenges of such a concept are the consideration of market behavior and the development of a concept for calibration/validation of complex spatial models. Further work is planned on the intriguing question what the relative effect will be of measures and scenarios on the total performance of the land-use & transport system.

A reliable integrated land-use & transport model contributes to the formulation of future policies in the fields of spatial planning and transportation. The theoretical challenge is to develop and validate a behavioral spatial model suitable to be implemented in a land-use & transport system for integrated analyzing of spatial and transport policy measures.
A detailed description of the content of the page is not possible due to the lack of visible text or images.
## Contents

1 INTRODUCTION
   1.1 Role of spatial planning .................................................. 1
   1.2 Land-use transport decision support system .......................... 2

2 PROBLEM SETTING AND GOAL
   2.1 Problem setting ......................................................... 4
   2.2 Aim of the research ...................................................... 4

3 THE SPATIAL SYSTEM .............................................................. 5

4 MODEL STRUCTURE
   4.1 Schematization of transport network and the spatial objects .... 8
   4.2 Actors in the model ....................................................... 9
   4.3 Settlement behavior modeling .......................................... 11
      4.3.1 Basic principles of the model algorithms ....................... 11
      4.3.2 Estimation of model parameters ................................ 12
   4.4 Time dimension .......................................................... 13
   4.5 Model application and testing ........................................ 14

5 OBSERVATIONS AND FURTHER DEVELOPMENTS ............................ 15
Contents

1 INTRODUCTION
   1.1 Role of mobile computing
   1.2 Design and performance evaluation

2 PROBLEM SETTING AND GOAL
   2.1 Problem setting
   2.2 Aim of the research

3 THE SPATIAL SYSTEM

4 MODEL STRUCTURE
   4.1 Components of transport network and the mobile service
   4.2 Accident location model
   4.3 Spatial interaction model
   4.4 Estimation of mobile communication

5 OBSERVATIONS AND FURTHER DEVELOPMENTS
1 Introduction

This paper discusses the research objectives of my dissertation research project. The dissertation study has started in March 1999 and the plan is to finalize the study in 2003. The dissertation study continues a line of research in the area of integrated land-use and transport modeling. A result of previous research in this area is the SPIRIT decision support system for an integrated analysis of policies in the field of land-use and transportation.

In this introduction chapter the role of spatial planning will be described and the SPIRIT system will be introduced. Other subjects of the paper are problem setting and objective, a description of the spatial system and a concept for spatial modeling. The status of the paper is preliminary according to the status of the study.

1.1 Role of spatial planning

An increasing importance is being given to a central and coordinating role for spatial planning. This central role of spatial planning and the associated basic need for coordination is illustrated in the flow diagram in Figure 1. The functions/activities and their spatial outlay as defined in the spatial plan form the basis for sector analyses in which balances are made between demands and available resources. Investments are planned and implemented in response to the demands for transport, water supply, housing, agriculture, industry and services, environmental protection, etc. Each of the sectors has its own methods of analysis and makes a certain use of space and characterizes this space on various attributes. Inter-sector relationships and/or equilibrium situations have to be taken into account. Based on the sector plans an overall evaluation can be made of the use of space. Based on this information a review can be made of the spatial plan leading to adjustments. Ideally the planning thus has a cyclical character in which information is exchanged and coordination takes place in both directions (directives from the spatial plan and feedback of information to this plan). The challenge in this planning cycle is to introduce the spatial planning options as much as possible from the start into the sector analyses.
It has become an accepted viewpoint that policy measures in the fields of housing, employment and adaptations in the provision and use of infrastructure, have impacts on spatial settlement as well as on the performance of the transportation network. Models addressing such integration have developed slowly, and transport policies are often still evaluated in its own terms. Due to environmental concerns, increasing congestion and declining suitable land resources the call for considering the system as a whole is growing rapidly. The SPIRIT research program of the last two years follows this trend and has resulted in a decision support system (DSS) including the land use system as well as the transport system. The DSS has been developed in a cooperation between Delft University of Technology and Demis BV (Delft, The Netherlands).

In most of the operational integrated land-use & transport systems the attention is focused on the transport system. The spatial system is sometimes simulated by no more than the accessibility parameter. Another spatial simplification is just implementing all the government plans. This method overestimates the guiding capabilities of the government and excludes the functioning of the land market.
The SPIRIT decision support system for integrated land use and transport planning focuses on a balanced approach including an operational spatial model. The DSS aims a dynamic modeling of the interactions between the spatial location of activities and transport, in order to generate insight into the effects of a large number of policy measures under different scenarios. The framework of the existing DSS can be divided into four modules, see figure 2.

For a description of the SPIRIT modeling concept and the structure of the analysis I like to refer to the paper "Decision support system for physical planning" (Verhaeghe 1998). The link between the SPIRIT research and the present dissertation research is that a newly developed spatial model will be implemented and tested in the SPIRIT decision support environment.

Figure 2: DSS modules and their connection

TRAIL Research School, December 1999
2 Problem setting and goal

2.1 Problem setting

A world wide growing population and intensifying activities results in a scarcity on resources in most sectors (water, land, transport). This is especially the case in the fertile delta regions where 70 per cent of the world population settles. The planning in these regions focuses on the conservation and increase of the resources and on the management of the demand. The evaluation of the performance of a region should incorporate supply-oriented measurements as well as demand oriented measures. Demand oriented measurements include measures influencing the levels and patterns of the demand. For a long-term plan the spatial development of the study region has to be simulated to generate the future spatial distribution of the type and level of functions.

Despite the size of the problems and the research efforts of the last thirty years in the field of integrated land-use and transport modeling the number of integrated models in practical-use is very limited. Another important weakness is the lack of a calibration concept for complex land use and transport models. Changes in land-use have a complex evolvement over time, frequently disturbed by many 'external' factors. Traditional statistical methods do not seem sufficient for the calibration of such a non-stationary system.

2.2 Aim of the research

The development of a spatial model which enables an integral analysis of spatial and infrastructure policies. The spatial model incorporates the land market, role of the government and the preferences of the land demanding functions. The model will be calibrated and validated in combination with the existing SPIRIT transport model.
3 The spatial system

This paragraph contains a short description of the spatial system. The description and understanding of the spatial system will be a continue part of study. New insights in the system will results in new functional specifications for the model. Key elements in the description are the consideration of scale levels, the involvement of various actors and their behavior, a sufficient separation of autonomous versus guided development and the regulatory and market situation.

Type of development
In the analysis of the development of a region a differentiation can be made between autonomous and guided development. "Autonomous" refers to the development of the region as result of the working of the free markets and "Guided" refers to steering of the development through governmental planning and regulations. An actual situation will usually be a mix of the two. The level of guidance by the government depends on the political involvement and the available regulatory apparatus to guide. The situation will be widely different for different countries. The spatial development in the Netherlands can be considered as strongly guided in the international comparison.

Scale levels and actors (figure 3)
Practically the spatial planning can be divided into three scale levels, the national planning, regional (provincial) planning and the township planning. Each of the levels has its own institutions and planning characteristics. The role of the governmental institutions differs greatly between countries as described above. A less dominant role for the government results in a stronger position for private parties as real estate developers and land speculators. These private parties will follow the preferences of their customers (the land demanding functions) and the most valued function will settle first. The role of the government in guided circumstances is to realize a spatial development according to the national or regional targets.

A totally guided spatial development is supply dominated. The government determines the choice set of the functions, in the most limited form there is only one option. The completely guided situation, according to the government targets, is undermined by the following processes: The land demanding functions are putting pressure on the government to change the policy or the functions rejected the planned locations and are capable to find alternative locations. This process is illustrated by the feedback loop in the figure.

The hierarchical sequence of targets from national targets, regional targets to local targets is not uniform. Governmental institutions are in competition
to meet their own targets. This competition occurs between different scale levels but also occurs within a scale level. For example, a national target for a balanced spatial distribution of social classes is in conflict with municipality targets to attract the high-income group. Another example is the competition between regions to attract the settlement of new companies.

Figure 3: Overview of spatial planning

Markets
The supply of available land and the land demanding functions meet each other in the land market. Actors in this land market are the government, real estate developers, and land demanding functions. As previously described, the government and the real estate developers regulate the level and kind of supply. The settlement choice of the land demanding functions is based on the preferences of the functions. The preferences of

Integrated spatial & transport modeling
the functions can be represented by utility modeling (see model structure). The government can steer the development in the desired direction by regulating the supply side or by measures influencing the preferences. The spatial model is developed to evaluate the impacts of infrastructure measures. This assumption has an impact on the level of detail of the simulation of the various markets.

The land market and the transport market are the central markets for the model target. Less important are related markets as the housing, labor or services market.
4 Model structure

The total framework of the DSS consists of several modules addressing strategy formulation, spatial allocation, transportation and evaluation. The present research focuses on the spatial module. In this paragraph a first set up will be discussed for the model structure of the spatial module. The existing DSS (SPIRIT) environment allows an efficient testing of different concepts. Basic elements of the model structure are the spatial schematization, institutional and market relations, number of land functions, time dimension and settlement behavior, those are described below.

A determining factor in setting up the model is the kind of analysis the model should be capable to address. An important division can be made into two types of analysis:

- Scanning various desirable future spatial layouts. Based on attractiveness the projected functions will be allocated to the unoccupied (agricultural, natural) land. The attractiveness differs for various scenarios of the political targets.
- Evaluation of various sets of measurements by simulating the spatial development based on the preferences of the functions and the reaction on the governmental measures. This type of analysis simulates the spatial development in a dynamic way.

Main difference between type a and b is that type b includes the choice behavior of the land users and their reaction on the government measures. A model capable of analysis type b is far more complicated than an instrument for the scanning (type a). For the goal of this study it is necessary to simulate the spatial development as in type b. Only this type can result in a for sector planning valid spatial pattern and level of settlement of the activities (e.g. demand of water or transport). The main challenge of type b is to identify the many uncertainty factors, their interrelations and the impacts on the results.

Another key specification for the modeling is the choice for a static or dynamic description of the functions. A lifecycle type of description can be used to address the dynamics of the urban functions. A dynamic approach makes it possible to simulate urban processes as deterioration or revitalizing of urban functions.

4.1 Schematization of transport network and the spatial objects

The project area will normally be part of a larger area with which it interacts. This interaction has to be preserved by appropriate boundaries.
In the SPIRIT system three zones are recognized with associated projections of population or economic activity.

- outer zone: the flows established for the base year are increased with an overall (national) yearly trend
- surrounding zone: the settlement in this zone which is clustered in some origin/destination nodes are increased with a trend which may be similar to the national trend or reflect regional difference
- project area: autonomous projections and/or specific measures (large housing complexes or industrial locations) are considered

The transport system has the same distinction and uses the principle that far distance transport is transported through national roads and that closer transport uses also regional roads.

In the project area the spatial schematization (using spatial cells) and the schematization of associated networks should be appropriately adjusted to each other. The size and shape of the spatial cells depends on the level of detail required by the particular study objective, the available data and requirements of the type of modeling. The size and shapes of the spatial cells should be arbitrary to allow the use of administrative units as spatial units. The model should be flexible to handle different kinds of spatial schematization.

4.2 Actors in the model

The two main types of actors in the spatial model are the government institutions and the land demanding functions. The real estate developers and agents can be considered as reacting on the settlement preferences within the regulations (zoning) of the government.

It seems arbitrary to model the future role of the government and the reaction of the government on the feedback information of the preferences of the land demanding functions. A more realistic approach is to illustrate the governmental options by using different policy scenarios (incl. level of guiding, resources).

The land demanding functions group consists of many different types of land-use functions, examples are agricultural, natural, commercial, institutional and residential. All of these functions can be subdivided into more homogenous categories. In this research the residential group will be studied in more detail. The residential group will be subdivided into various categories based on social-economic characteristics (e.g. income, household size). The number and size of the categories will be determined by the available data for a region and by requirements of the settlement behavior modeling method.
Spatial and transport system (t0) 
Features of the site and spatial relations

- Deteriorated functions
- Existing Land Reservations

Zoning policy

- Number and kind of functions
- Features of functions

Expert judgement

Estimation of the preferences of the functions and the utility functions (incl. Willingness to pay of the functions)

Available sites per function (t1)

- Features of the site
- Spatial relations

Spatial measures

Infrastructure measures

Spatial measure ments

Attractivity of sites for the different functions

If a site is overcrowded
Then land price goes up

Allocation of the functions

Spatial distribution of the functions and transport demand (t1)

Figure 4: concept for spatial modeling
4.3 Settlement behavior modeling

User behavior modeling is quite well developed for transportation behavior but less for settlement behavior. A literature study of the application of discrete choice models in the transport sector can function as a starting point for the use of discrete choice techniques in spatial modeling. The decision-makers in the discrete choice modeling are described under actors and the spatial zones can be considered as the alternatives. The governmental zoning policy determines the available alternatives (spatial zones) per decision-maker (land demanding function).

Earlier attempts of using discrete choice modeling in land use models are or not operational or the models are unbalanced with a domination of the transport aspects. The use of only the transport system as explanatory variable for spatial development is way too limited, especially in areas with a high-density transport system (like the Netherlands).

The attributes of the alternatives in the spatial planning can be ordered in three groups:
- physical characteristics of the spatial zone
- surrounding of the zone, attractive or unattractive for function
- spatial relationships of functions by using the transport network (incl. performance transport system)

A spatial database has to be built up containing the physical and land-use attributes. The accessibility of a zone for a specific function is the result of the performance of the network (transport model), the spatial distribution of functions and the preferences of the function. The large database can be collected and an existing transport model can be used to calculate the transport system performance. The spatial model and the transport model can exchange the data in time, the spatial model uses the results of the transport model of the previous time step, or the spatial and transport model can exchange their data in an iterative process.

4.3.1 Basic principles of the model algorithms

Presently the model specifications are not defined in detail but some of the basic principles can be presented. The central principle of the allocation module is that a function settles down in the most attractive zone for the function. Limiting conditions are the zoning policy of the government or the case if two different functions want to settle down in the same location. In that case the most valued function settle first (highest willingness to pay).

The attractiveness of a zone for a function is the combination of location characteristics, as features of the zone, the surrounding zones, spatial relations, and the preferences of a function. Utility functions are describing the attractiveness for the different functions. The missing information and
main uncertainty in these utility functions are the preferences of the land demanding functions (see estimation of model parameters).

The spatial relations also include the spatial distribution of for example jobs or housing supply. The spatial dimension of the supply-demand markets can be addressed by a kind of gravity principle which is already for decades very familiar in spatial modeling (see Lowry). The gravity parameter will be unique for every land demanding function based on their activity pattern and preferences. The spatial distance can be calculated by the transport model, this opens the option to include road congestion or road pricing in the distance parameter.

The land prices and the willingness to pay of functions play an important role in the function transition of a site. The most important parameters to estimate the land prices are: the kind of activity, the life cycle of the activity and the accessibility. The willingness to pay of a function is based on the characteristics and preferences of a function. The principle that the most valued function settles first has to be operated. A way is that if a location is the most attractive for various functions then the land price of the location will be raised (by a small step). In an iterative process the land prices of the different ‘overcrowded’ locations are raised until each function is allocated to an original location

### 4.3.2 Estimation of model parameters

Improved physical and land-use data collection methods result in a reliable spatial database. As described above the preferences of the land functions are the main uncertainty factor in the utility functions. The utility functions include parameters as type of houses, recreation area, plot size per household. The preferences (utility) of the functions contain a systematic and a random part.

The preferences of the functions can be estimated by using different types of information:

- **Analysis of historical time series:**
  In this analysis the preferences of the functions are gathered with statistical techniques. Assumption is that the preferences are unchanged in time. This type of analysis can be used in areas where the preferences of functions are the only driving force of the unguided allocation (Zondag B., 1997). But it seems to have little potency in areas with an unclear government regulated and free allocation. Another disadvantage of this type of analysis is the necessity of large historical data sets and the sensitivity of this time series analysis for disturbances in the data.

- **Stated preferences (SP)**
  By using the stated preference survey technique individuals (out the land demanding functions) are asked about what they would prefer in various hypothetical situations. A very basic problem with SP data collection is
how much faith can we put on individuals actually doing what they stated they would do when the case arises (Ortuzar, 1994). Recent stated preference methods have improved the results. But these methods are very demanding in their requirements for trained survey staff and quality-assurance procedures.

- Expert judgement

Using the knowledge of experts can be an alternative way to gather the information about the preferences of the functions. In this case not the residents themselves are asked what they prefer but project developers or brokers who are expert in the settlement preferences of the various social-economic population groups are asked to choose in hypothetical situations.

Practical limitations have a strong influence in determining the most appropriate type of survey for a given situation (Ortuzar, Willumsen 1994). In this PhD research stated preference data collection doesn’t fit within the financial and time budget. Expert judgement of will be used to estimate the residential settlement preferences. The nature of the expert judgement data and the objective of the model will be determinant factors in the choice of model estimation techniques.

4.4 Time dimension

The new spatial distribution of the functions results directly in a new spatial pattern of the transport demand. The effects of spatial changes are directly visible in the transport system performance. On the other hand the effects of changes in the performance of the transport system can be considered as long term effects on the spatial system. The simulation time of models illustrating these reverse effects has to be a long-term (strategic) period.

The following processes influencing the spatial system can be considered as time dependent:
- economical or physical deterioration of functions (lifecycle)
- allocation of new functions
- changes in the behavior of the functions

The model has to be dynamic to include the time dependent spatial processes and to use the results of the transport model. The time step size can be flexible up to:
- type of region (high or low migration or economical growth figures)
- computation time of the application
- planning period and objective
4.5 Model application and testing

An in the present study newly developed spatial model will be implemented in the decision support environment to allow an efficient testing and to facilitate the evaluation of various measures under different scenarios. The model should facilitate the evaluation of the impacts of transport measures on the long-term spatial distribution of functions. Vice versa the DSS can evaluate the impacts of the spatial and transportation measures on the performance of the transportation policy. Multi criteria analysis or cost-benefit techniques can be used to compare the various sets of measures.

Besides the options for policy analysis the model has to be tested on the quality of the reproduction of the actual system. The spatial model will be tested in real world case studies, the idea is to set-up two case studies under different circumstances (one relatively guided and one relatively unguided). The model has to be tested on a sufficient spatial representation as well as its behavior over time.

The many parameters of an integrated spatial & transport model makes the calibration of such a system a complex activity. Concepts for the calibration of complex and non-stationary systems have to be studied. A sensitivity analysis has to trace the main uncertainties and their impacts on the result.
5 Observations and Further Developments

The proposed modeling concept in this paper focuses strongly on the settlement demand; what are the settlement preferences of the functions and how can the government influence these preferences by spatial or infrastructure measurements in the direction of the desired societal targets. These questions can only be analyzed when the model is implemented in a decision support environment with strategy formulation, evaluation, visualization and transport modeling facilities. Key elements are mapping of spatial characteristics, spatial and network schematization, and an efficient attachment of application modules to this data and networks.

The approach of the spatial modeling can be considered as an 'engineering approach': focus on the spatial and infrastructure measures and their main impacts on the physical flows and the allocation process. The approach focuses less on the measurements or impacts of the social-economic markets (e.g. housing, labor). The detail level of these markets in the model depends on the necessary level for a realistic evaluation of spatial and infrastructure policies.

Further subjects of study in the research project will be:
- efficient handling of spatial data, guidelines for a spatial data base
- settlement behavior modeling, estimation of preferences
- identify the main uncertainties and their impacts on the results
- concept for calibration of integrated spatial & transport modeling
- case study, including policy evaluation
References

- Miyamoto K. and Kitazume K., A land-use model based on random utility / rent-bidding analysis (RURBAN).
- Miyamoto K. and Udomsri R., An analysis system for integrated policy measures regarding land-use, transport and the environment in a metropolis.
• Verhaeghe R.J., Zondag B., *Decision support system for physical planning, integration of land-use and transportation planning*, Delft 1998.


References

Policy Making in the Netherlands:

the limits of an open approach

TRAIL Research School, Delft, December 1999

Author
Ms. Drs. M. Esselbrugge
Department of Public Administration, Erasmus University Rotterdam

Thesis supervisor
Prof. Dr. Ing. G.R. Teisman
Department of Public Administration, Erasmus University Rotterdam
Contents

The paper presents the preliminary results of a Ph. D project aimed at developing a dynamic process architecture (design + management) for complex decision-making within large-scale infrastructure projects. Some key words in this study are democracy, ecology, partnerships, interactive policy making and process management.

Abstract

1. Introduction ............................................................................................................................................. 1
2. Tensions between open and closed approaches .................................................................................... 3
3. Typology of open and closed approaches ............................................................................................... 6
4. Causes of a closed approach ......................... 7
6. Appreciative Management .................................................................................................................. 12
7. Final remarks ...................................................................................................................................... 17

References .................................................................................................................................................. 19
Contents

1. Introduction

2. Tropisms between prey and predator adaptations

3. Techniques of clean and clean digestion

4. Course of a clean application

5. Case of data - SPECIFIC HISTORY

6. Administrative Management

7. Plan leverage

8. References
Abstract
This paper presents the preliminary results of a Ph. D project aimed at developing a
dynamic process architecture (design + management) for complex decision-making
processes within large-scale infrastructure projects. Some key words in this study are
mobility, ecology, partnerships, interactive policy making and process management.

In search of better and more efficient policy-making processes for large-scale
infrastructure projects it has been recognised that government is no longer able to
single-handedly develop and implement policies or decisions without consulting
citizens, pressure and environmental groups or private parties. Therefore, the Dutch
national government is redesigning decision-making processes in a more open and
interactive way. The common thought is that more open, participatory and interactive
methods of decision-making will be able to cope with a variety of actors and their
often-conflicting opinions, preferences and claims. However, these open approaches
seem to show some tensions.

This paper will attempt to give an answer on the question: Are the open decision­
making processes able to deal with (groups of) actors and their often-conflicting
preferences in an adequate way? And does this mean that closedness will be history,
or maybe also represents an important value. When the latest is true, and in this
contribution it is stated, than a new topical subject in public administration originated.

Decision-making is open and closed simultaneously.

This is the reason that a pluricentric analysis model has been developed that also
takes into account the role of non-participating 'thirds', i.e., third parties, perceptions,
resources or moments within and outside the process. Based on this model we
introduce the management technique known as appreciative management. The
most important assumption underlying appreciative management is that in every
process something works, and change can be managed through the identification of
what works. This is virtually the opposite of the approach taken by traditional
problem-solving techniques.
The project aims to develop and implement an efficient and effective software application for monitoring and managing greenhouse gas emissions. The system will be used to track and report emissions data for various projects and activities. The software will integrate with existing systems and provide real-time data analysis.

The project will involve the development of a new database management system, which will be used to store and analyze the emissions data. The system will be designed to be user-friendly and will provide reporting tools for stakeholders.

The project team will consist of experts in software development, database management, and environmental monitoring. The project will be managed using Agile methodologies to ensure flexibility and adaptability to changing requirements.

The project timeline is divided into several phases, each with specific deliverables and milestones. The project is expected to be completed within the next 12 months, with regular progress updates to stakeholders.

The project will have a significant impact on reducing greenhouse gas emissions, improving environmental sustainability, and enhancing the reputation of the organization involved.
1. **Introduction**

Decision-making about the need and necessity of infrastructure projects is subject of criticism in several European countries. The discussion is no longer only evolving around the question of which projects should be realised. The question how these processes should be organised and managed is also part of the discussion. The problems related to decision-making about infrastructure issues are well-known. Firstly, most projects are very expensive and government is blamed for being incapable of working efficiently. Secondly, it takes a long time to make decisions about infrastructure investments. Finally, criticism focuses on the inability of (national) government to deal with groups of citizens in an adequate way (Teisman, 1997b). Governments find themselves struggling between the diverging preferences of society. Infrastructure policies not only have to meet the claims of 'homo mobilis' but also must ensure a safe and clean environment. To cope with these high demands of society, the Dutch National government is redesigning the decision-making processes in a more open and interactive way. New actors are invited to participate, like private organisations, citizens-groups and individuals. It is now 'commonplace that the democratic process needs to be opened up to involve wider sections of the populace more directly in decision-making' (Cochrane, 1986:51). This is also considered desirable from a democratic point of view because policies or policy projects should be linked to a variety of existing preferences in society (Teisman, 1997a). Interactive decision-making, public private partnerships and creative competition are examples of more open approaches.

In the Netherlands it is recognised that government is no longer the central governing authority in society. Rather than depending on traditional hierarchical forms of organisation, public managers are experimenting with flatter structures and more participatory ways of organising. The open approaches of policy-making are aimed at involving citizens, private parties and societal groups in the early stages of policy development, where they can have input into the problem-defining process and the development of alternatives. In contrast, in traditional more closed policy-making processes the doors are not open to citizens, societal groups or private parties until after a decision has been made or a government memorandum drawn up. Government, nowadays, has to deal with different categories of actors and therefore a more transparent style of decision-making is demanded. It is assumed that policy making of any significance should be the result of interactions between public and private partners.
Policy- or decision-making becomes a process in which different preferences and several actors have to be interwined to reach an acceptable compromise. The common thought is that more open, participatory and interactive methods of decision-making will be able to cope with a variety of actors and their often-conflicting opinions, preferences and claims. However, these open approaches of decision-making seem to show some tensions.

In this paper it will be stated that both open and closed approaches are valuable in the policy-making process. The paper will also offer you a typology of open and closed approaches, illustrated in the case A4 Delft – Schiedam National Highway. Furthermore, it will explore the causes of closedness in decision-making processes and introduce the management technique known as appreciative management. Appreciative management is not just a technique for problem solving but a process for change, which will become far more important for managing openness and closedness in policy-making processes.
2. Tensions between open and closed approaches

It became clear that the rationality of central government was limited and that it could not control society itself. Therefore, the traditional, more closed, decision-making processes did not seem very suitable in a context in which there were more participants involved, with their own views of reality, and none with enough steering ability to determine the strategies of the others.

Characteristics of a more closed policy style are a single, defined problem-formulation, anchored in one-sided, formulated aims (Bekkers, 1996). In a network society, such as the Netherlands, in the field of infrastructure policies, there is a growing need to steer on the basis of a common conception throughout a collective policy practise. This requires a more open policy style, which is based on the assumption of plurality. Steering or governing is seen as an interaction process between a diversity of actors, from governmental agencies to individual citizens. This is why the government decided on a more open, participatory and interactive way to organise decision-making. Thus, openness is seen as a reaction to a closed approach. At the same time, it should be recognised that other actors, like private organisations, interests groups and individual citizens are also faced with considerable limitations, particularly in terms of resources and instruments at their disposal and their cognitive competence. These more open and interactive ways have their limitations. This should be considered whenever a process of decision-making is designed in an interactive way.

The tendency towards more openness in the decision-making process generates some tensions which can be explained in terms of an open and a closed approach. In this section four fields of tensions between an open and a closed approach will be distinguished.

Firstly, we can observe a tension in the field of actors. Openness in this field implies that all relevant actors will be involved in the process of decision-making (de Bruijn e.a., 11998). This is necessary because otherwise, when certain actors are excluded, they will try to block the process. Or more positively formulated, a variety of actors is preferable because these actors have certain resources at their disposal, as also a specific expertise necessary to come to a solution. However, by doing so the number of actors will increase. The involved actors all have their own ideas about the problem and possible solution(s). This will lead to a more complex process, which result in uncertainty about the outcome(s) of the process. It also seems that a closed approach represents an important value in this field.
Secondly, we can distinguish a tension in the field of perceptions, which refers to the inclusion of a variety of opinions and interpretations. This is advisable from a democratic point of view, due to the fact that in a democratic society we should (re)consider all opinions and interpretations which are existing in society. Also, openness in the field of perceptions is desirable to reach a higher quality with respect to the content of the process (de Bruijn e.a., 1998; Teisman, 1997a). However, continuous openness in this field also creates risks. It can lead to a confusion of tongues, impractical ideas and a multitude of perceptions can result in not seeing what really counts. Again, a closed approach seems to be preferable under certain conditions.

Thirdly, there is a tension in the field of information, which is a very important resource for the decision-making process. Openness in the field of information is needed to come to well-considered decisions. Actors have access to different kinds of information and most of this information should be considered in solving a problem adequate. The confrontation between these different kinds of information can increase the quality of the information used. However, openness can lead to an overload, or again to incomplete information and thus to non transparency. This means that the actors involved in the process cannot deal with the huge and divers amount of information: they cannot distinguish relevant information from details of minor importance. Nevertheless, decisions have to be made. In such situation the process should focus on reducing the risks which are at stake by making decisions. This means that a closed approach is necessary at some time during the process.

The last field of tension between openness and closedness can be found in the field of decision-making moments. This field refers to the coupling between the several decisions made during the process and between the final outcome of the process and the activities following. For example, the interactive moment in the process linked with the formal political decision-making moment. In the Netherlands the outcome of an interactive process is still an informal document, as the formal decision is still taken by political bodies. Tensions between this interactive form of decision making and the formal procedures can easily appear. It seems necessary to give a clear description about how (part of) these decisions will be coupled, as also and the (legal) status of the outcome of open, interactive processes.
These tensions illustrate that both openness and closedness are valuable for the decision-making process. Because, continuous openness does have its risks: confusion of tongues, slackening, unrealisable ideas, et cetera. The possibility exists that interactivity will only lead towards inertia. Therefore, we should recognise that a closed approach within the process also has value. Considering, only a limited number of actors, preferences and means, implies that a single, unambiguous, problem definition can be given. The question that need to be answered is, 'When is the open approach preferable to the closed approach in the policy-making process and visa versa?'.

TRAIL Research School, December 1999
3. **Typology of open and closed approaches**

Referring to the four fields of tensions between open and closed approaches, it can be imagined that a decision-making process is open and closed simultaneously. If a decision-making process is open and closed simultaneity than a new theme in public administration originated: the question of open and closed approaches. What are the consequences of organising (process-architecture) and managing (process-management) decision-making according to the simultaneous inclusion of open and closed approaches in a decision-making process?

To get a better understanding about the concept of openness we need to reflect on the concept of a closed approach. Network theorists are using a wide variety of interpretations of the concept ‘a closed approach’. However, a number of questions are still left unanswered, for example: What causes a closed approach and how can it be recognised in a (traditional) decision-making process?

The closed approach exists in the four distinguished fields were tensions can occur. Those four fields are based on the questions: who (actors), what (perceptions), how (information) and when (moments). ‘Who’, ‘what’, ‘how’ and ‘when’ are aspects of interactions. Hereby following Lasswells’ ‘who gets what, how and when?’ (1958). We can then speak about a closed approach when the decision-making process is restricted for actors who wants to participate, notably when: some perceptions are not debatable, if not all relevant information is freely exchangeable and if the coupling of decision-moments is predetermined (see table 1). Therefore, open and closed approaches are pluri-dimensional concepts.

![Table 1: Typology of open and closed approaches](image)

---

**Policy Making in the Netherlands**
4. Causes of a closed approach

In this section a theoretical examination of the closed approach will be explored. In the literature of networks we can find several insights into the closed approach and policy networks (e.g. Rhodes, 1980; Jordan, 1990a; Schaap and van Twist, 1997; De Bruijn and Ten Heuvelhof 1991, 1998). In keeping with Schaap and van Twist we can distinguish the closed approach of separate actors and, what they call ‘closedness’ of networks or processes. In other words, there is an internal and external perspective to the conceptualisation of closedness. Distinguish an internal and external perspective. The internal perspective relates to the view of the individual actor (within the actor) and the external perspective may be seen in the process or policy-network (within the process or network). It may be possible that an individual decides not to co-operate or to participate, because he or she does not want to commit himself to the outcome of the process (internal). It is also possible that the individual is willing to participate but excluded from the process because the actors already involved want to keep such complex processes manageable (external).

Closedness can be the result of a conscious strategy or of an unconsciously applied rule. For instance, in the field of actors the exclusion of actors can be caused by not knowing the jargon. Regarding the benefit- and-necessity discussions by decision-making processes of major infrastructural projects you see the exclusion of local people due to the jargon pragmatic civil servants use in terms of noise standards and risk contours. This results in a unconsciously extern closedness in the field of actors caused by the culture of the policy network or process. A conscious strategy from an internal perspective is present when actors use their power. For instance, the actor is not prepared to provide another actor with resources and therefore refuses to interact. Within this example you can discover the intertwining of different types of closedness: the conscious internal closedness in the field of information (resource) causes the conscious internal closedness in the field of actors. An example of the latter is for instance that the actor finds the transactions costs to high and therefore exclude himself from the process, he is not willing to interact.

We can now represent schematically (see table 2) possible causes of a closed approach, which can serve as a basis for organising and managing complex decision-making processes. It is not an attempt to provide a full description.
Table 2: A pluricentric policy analysis model

<table>
<thead>
<tr>
<th>Perspective</th>
<th>Dimension</th>
<th>Aspects of interaction</th>
<th>'Who'</th>
<th>'What'</th>
<th>'When'</th>
<th>'How'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra (within actors)</td>
<td>Conscious</td>
<td>Free rider behaviour</td>
<td>(Ir)rationality</td>
<td>Individual</td>
<td>Power-oriented strategy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(ban certain points of view)</td>
<td>Veto power (refuse to interact)</td>
<td>strategy</td>
<td></td>
</tr>
<tr>
<td>Unconscious</td>
<td>Bias</td>
<td>Frame of reference</td>
<td>Self-interpretative</td>
<td>Veto power</td>
<td>Strategic co-operation</td>
<td></td>
</tr>
<tr>
<td>Internal (within the process)</td>
<td>Conscious</td>
<td>Formally arranged</td>
<td>A particular issue is declared to be out of order for strategic reasons</td>
<td>Veto power</td>
<td>Strategic co-operation</td>
<td></td>
</tr>
<tr>
<td>Unconscious</td>
<td>Informal rules of behaviour (culture)</td>
<td>Group-think processes</td>
<td>Self-referential</td>
<td>Failure to appreciate the resources</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An examination of the closed approach will cover both theory and practise. In the next section an examination of a closed approach is presented in an empirical case; the process of decision-making by the Delft-Schiedam National Highway.
5. Case: A4 Delft –Schiedam National Highway

The decision-making process of the A4 Delft-Schiedam National Highway in the Netherlands has a long history, which started in the late 1950s. In those days, the first plans for a second connection between the Hague and Rotterdam were drawn up. Already in 1965 the Dutch Cabinet had decided to construct the highway, but even now the highway is still not built. In this process, it seems impossible for all of the actors to look in the same direction. This does not seem very surprising, because the highway would sever the “Midden Delfland” countryside and cuts into two residential areas in the cities Vlaardingen en Schiedam. Environmental values, mobility and liveability are playing a major role in this decision-making process. Until now, this (value-) pluriformity has ended in a long chaotic process. It is important to see how processes like this are advanced and how they can be improved to meet the high demands of society in the 21st century.

The main argument to construct this road was the congestion-problematic on the A13. Since the mid-seventies environmental and pressure groups have challenged this position, especially the pressure group known as Stop Rijksweg 19. Until than nobody had questioned the need for the development of the A4Delft-Schiedam. For the first time Rijkswaterstaat (engineering department of the Ministry of Transport, Public Works and Water Management) was confronted with diverse and conflicting perspectives. Instead of working with Stop RW 19, the RWS decided to consciously exclude this group and in that way exclude their perspective that was and still is opposed to the A4. But during the seventies, at the same time as the A4 discussion was taking place, Parliament debated the first policy document on spatial planning that considered ecology and the environment to be important issues. It was in this arena that the pressure group got a hearing and their protest ended with the passing of the ‘Voortman motion’, which asked for postponement of the construction activities of the A4 and an investigation into alternative solutions. Because the decision-making process involving the A4 had been closed to the pressure group they felt they were being confronted with a fait accompli. This is why they tried to influence the process through other channels. Stop RW 19 celebrated their first success when the Voortman motion was accepted.

During the eighties the word ‘inpasbaarheid’ (meaning fitting in: that the infrastructure should fit within the surrounding environment) seemed to be the new catchword. This led to the debate becoming very technical with the RWS the central actor in the process. Finally, in 1988 consensus was achieved about construction of the A4 under specific ‘fitting in’ conditions. In the beginning of the nineties construction activities started again, and at the same time the pressure group restarted its protest campaign.
This time they used official juridical channels to oppose noise pollution. This time the RWS won. Even so, the lesson to be learned is that exclusion of the pressure group, which ended in a confrontation over details within official channels, led to their organising their opposition with great pent-up range.

European legislation caused another delay in the construction of the A4 Midden Delfland. This legislation, dating from 1987, made it compulsory to conduct an environmental impact study (MER in Dutch) for national highways in cases where the spatial planning had not yet been completed. Because of the diverse perceptions concerning the A4 Midden Delfland, the RWS decided to combine the MER proceedings with the legal procedure for determining where the roadway would be constructed and started the whole decision-making process over again. This time they chose for a joint and open process in which several relevant actors, including the pressure group Stop RW 19, participated. The process was open in regard to 'actors', but could be considered closed concerning 'perceptions'. This is because the discussion was dominated by the RWS who used technical jargon and focused the process mainly on the subjects of traffic and transport. In this arena participants could only speak within a traffic/transport framework.

Some issues, such as those concerning the environment, were declared only relevant as pre-conditions for the construction of the highway. The pressure group therefore had no influence at all.

It was no surprise that the process ended in the following position taken by the Cabinet (1996): 'The national highway A4 Midden Delfland should be constructed at ground level in Midden Delfland and between Schiedam and Vlaardingen on the already constructed foundation with a budget of 370 million Dutch guilders.' The other relevant actors did not share this position. Several alternatives were developed, some of them supported by private parties. But the Minister at the time, Mrs Jorritsma, rejected these because she found them much too expensive. A stalemate was born.

The process restarted, this time on a regional level and excluding the RWS. It ended with a letter signed by all relevant regional parties, with the exception of Schiedam. In this letter the assumption was made that construction of the highway was inevitable. A resolution seemed at hand, but then Parliament accepted a motion by Van Heemst (17-12-98) in which the budget for the construction of the A4 was incorporated into other infrastructure projects. It can be said that the parties involved with the A4 did not recognise the interference in their process and other infrastructure processes because they had become self-referential (internal, unconscious closedness concerned with the 'timing' aspect of decision making).
Presently, due to the limited availability of finances for infrastructure investments, the Ministry of Transport, Public Works and Water Management is exploring the possibility of developing and building the road in a public-private partnership. In actuality, they are asking for private finance only.

This brief description shows the relevance of a pluricentric policy analysis model. It sketches how a decision-making process develops in terms of openness and closedness. The aspects of interaction where openness or closedness can occur can function as leverage points for process management, such as appreciative management.
6. Appreciative Management

"Societies, networks, processes and organizations need less fixing, less problem solving and more positive images born of appreciation."

A pluricentric perspective on joint governance policy processes does not necessarily result in a pessimistic view of the possibility of influencing such policy processes. On the basis of the theoretical framework outlined in section three and four, it would appear to be quite possible to identify leverage points at which attempts to exert influence could be made. The theoretical framework has four aspects available for influencing actors (intra) and processes (internal). The leverage points for managing complex decision-making processes are the four distinguished aspects of interaction: actors, perceptions, resources and moments with respect to decision- or policy-making. These aspects can be influenced towards openness or closedness (or inclusion and exclusion) of thirds. Process management can than be viewed as an instrument to influence ‘who gets what, when and how’ (cf. Lasswell, 1959) and puts the emphasis upon non-represented ‘thirds’. This is of major importance because when the thirds are considered they have the ability to influence the development of the policy process and subsequently the policy content. The table below (table 3) outlines the four aspects at the actor and process levels and the corresponding activities. One of the techniques to achieve these activities is called appreciative management.

Table 3: Actor- and process management

<table>
<thead>
<tr>
<th>Actor management</th>
<th>Actors</th>
<th>Perceptions</th>
<th>Resources</th>
<th>Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selective activation</td>
<td>Selective</td>
<td>Changing perceptions</td>
<td>Mobilising resources</td>
<td>Changing powers</td>
</tr>
<tr>
<td>Changing relations between actors</td>
<td>Intertwining perceptions (consensus building)</td>
<td>Changing the distribution of resources</td>
<td>Anticipating decision rules</td>
<td></td>
</tr>
</tbody>
</table>

Introduction

Appreciative management is inspired by Appreciative Inquiry (AI). AI was developed by David Cooperrider as a new paradigm with the potential to replace the conventional problem-solving methods of organisation development. This approach seems obvious, we know from our own experience that we can look at what isn’t working and start problem-solving. If we do so, we focus on difficulties in the past, people become self defeating.
and feel that life is hopeless. When we ask them about their successes, they
become enthusiastic and start to hope again. In my opinion this method can
be translated to ‘changing environments’ in the field of public administration
as well, like partnerships and decision-making processes.

‘By definition, a problem implies that one already has knowledge of what
“should be;” thus one’s research is guided by an instrumental purpose tied to
what is already known. In this sense, problem solving tends to be inherently
conservative; as a form of research it tends to produce and reproduce a
universe of knowledge that remains sealed.’

Appreciative inquiry is a way of thinking, seeing and acting for powerful,
purposeful change in organisations (Hammond, 1999). The most important
assumption of AI is that in every organisation something works and change
can be managed through the identification of what works. Translated to
policy-making the assumption of AI is that in every joint governance process
something works and change can be managed through the identification of
what works. The traditional approach to change is to look for the problem,
diagnose it and find a solution. And since we look for problems, we tend to
find them. The AI method, on the other hand, looks at processes /
organisations with an appreciative eye.

Is it a very naïve approach? Not really. We are very good at talking about
what doesn’t work. It never occurs to us that we can ‘fix’ a process by doing
more of what works. The objective is to improve interactions in the process
while recognising that individuals from diverse cultures have different ways of
interpreting ‘reality’. When the appreciative approach is applied, participants
are asked to share examples from their own experience about a successful
policy-making process. Participants analyse the circumstances that make a
process successful and identify common themes. Instead of taking away a
list of don’ts, participants leave inspired to re-create those circumstances in
as many processes as possible.

<table>
<thead>
<tr>
<th>Problem Solving</th>
<th>Appreciative management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of problem</td>
<td>Appreciating and valuing the best of ‘what is’</td>
</tr>
<tr>
<td>Analysis of causes</td>
<td>Envisioning ‘What might be’</td>
</tr>
<tr>
<td>Analysis of possible solutions</td>
<td>Dialoguing ‘What should be’</td>
</tr>
<tr>
<td>Action planning</td>
<td>Innovating ‘What will be’</td>
</tr>
<tr>
<td>Basic assumption: a process is a problem to be solved</td>
<td>Basic assumption: a process is a mystery to be embraced</td>
</tr>
</tbody>
</table>

Adapted from Cooperrider and Srivastva (1987) ‘Appreciative Inquiry Into Organizational life’
in Research in Organizational Change and Development. Pasmore and Woodman (eds.) vol.
1, JAI Press.
The role of perceptions in change

Processes are made up of individuals who form configurations (or groups) to get work done. The configurations behave according to the rules of behaviour. Assumptions are the set of perceptions shared by a configuration that causes the configuration to think and act in certain ways. Those shared perceptions can, for example, consciously or unconsciously cause a closed approach towards ‘thirds’ (see section 4). The longer the perceptions are in effect, the harder it is for the configuration to accept any new information, actor or third perception that contradicts the belief. The shared set of perceptions of a group is a powerful force. One needs to understand what the assumptions are in order to predict how the diverse configurations in a process will act and how the process will progress. To understand appreciative management, one must understand its assumptions.

- In every society, organisation, process or group, something works;
- What we focus on becomes our reality;
- Reality is created in interaction, and there are multiple realities;
- The act of asking questions of an organisation, process or group influences the organisation, process or group in some way;
- People are more comfortable and have more confidence in journeying to the future (the unknown) when they carry forward parts of the past (the known);
- If we carry parts of the past forward, they should be what is best about the past;
- It is important to value differences;
- The language we use creates our reality.


In order to see data that conflict with our perceptions, we have to break outside of our filter or frame. Sabatier calls the frame deep core beliefs; Edgar Schein and Peter Senge call it a mental model. All point out that the frame can inhibit openness because data that does not fit the frame is inconceivable and is often not able to be discussed.

AI method

Breaking the frame or changing and intertwining perceptions can be achieved by using appreciative management, because the participants share their data by conducting interviews with the partners. In the processes of looking for and sharing what worked in another policy-making process, the ‘dysfunctional’ actor or configuration realises that they could work in similar ways and have similar results.
Pair interviews are recommended as the most effective tool for exploration. When the pairs finish the inquiry there is a great deal of information between them. The goal is to share that information with the larger group in order to uncover common process conditions that occur when a joint-governance process is successful. We want to uncover these conditions in order to know how to do more of what worked. The participants are asked to share the best story, or the most ‘quotable quote’.

The tangible result of the inquiry process is a series of statements that describe where the person or organisation(s) wants to be, based on the high moments of where they have been. Because these statements are grounded in real experience and history, people know how to repeat their success.

“To the empowering principle that people can withhold legitimacy, and thus change the world, we now add another. By deliberately changing the internal image of reality, people can change the world” (Willis Harman)

In short the method can be described as a 4D-cycle.

Discovery: Looking for what is working. Appreciating the best of our experience.
Dream: This is to consider what might be. Envisioning Results.
Design: What should be the ideal? Co-constructing
Destiny: How to empower, learn and adjust or improvise? Sustaining.

“The real act of discovery consists not in finding new lands but seeing with new eyes” (Marcel Proust)

An example
In order to develop a project called Essepark in a public-private partnership arrangement, the public and private participants were asked to explore process conditions, as mentioned above, for successful partnership. The questions were these:

Describe a successful complex partnership process from your own experience.
1. What conditions made this partnership successful?
2. What made this partnership unique?
3. What part of the partnership would you implement in future partnerships?
Story after story recounted partnerships in which people were enthusiastic about what was reached in partnership. Based on their own experience the participants mentioned the following process conditions that made the partnership successful:

- directness
- space for individual success
- the willingness to try to understand each other
- knowing one's own interests and goals
- respect and appreciation for others in the process
- offering acknowledgement
- showing vulnerability
- having a good team
- having an excellent start to the process

"Playing with PPP", game simulation commissioned by RWS/AW 1-2 July 1999 in cooperation with Awareness. Participants were i.e. Paul Baks and Piet Hemmen.

Appreciative management is not just a technique within the problem-solving paradigm. Realise that the process is just that – a process. It lives and change will occur within it. There is no end, because it is generative.

"The key point is that all of our cognitive capacities – perception, memory, learning – are cued and shaped by the images projected through our expectancies. We see what our imaginative horizon allows us to see. And because “seeing is believing”, our acts often take a whole new tone and character depending on the strength, vitality, and force of a given image" (David L. Cooperrider).
7. Final remarks

This paper presented the preliminary results of a Ph.D study aimed at developing a dynamic process architecture for complex decision-making processes within large-scale infrastructure projects. A final dynamic process architecture will only be drawn up following an analysis that includes case studies.

In search of better and more efficient policy-making processes the Dutch government has opened up the decision-making process to involve societal groups, citizens and private parties. In these so-called open approaches it is important to make the distinction between openness and closedness on all four aspects of interaction. When analysing a complex decision-making process or developing a process-architecture, the distinction between an open approach and a closed approach should be at the heart of the model. It is, according to Lasswell, a question of power: 'Who gets what, when and how?'

Process management can be viewed as an instrument to influence 'who gets what, when and how?' One of the techniques to achieve this is called appreciative management. A technique that, in my opinion, can be more fruitful than traditional problem-solving techniques.
For information please contact:

Ms. M. Esselbrugge (Monique)
Erasmus University – Department of Public Administration
P.O. Box 1738, NL-3000 DR Rotterdam
Fax : (+31) 10 408 9099
Email : Esselbrugge@bsk.fsw.eur.nl

For more information on Appreciative Inquiry see the following websites:

http://www.appreciative-inquiry.org
http://www.serve.com/taos/
References


• Esselbrugge, M., (1999), 'Besluitvorming A4 nader bezien', tussenrapport; Erasmus Universiteit: Rotterdam.


• San Francisco, CA: Jossey-Bass Inc.
The Role of Transport Costs in Goods Transport

An analytical framework

Authors
Drs. H.A.C. Runhaar
Dr. B. Kuipers
Faculty of Technology, Policy and Management, Delft University of Technology

Thesis supervisor
Prof. dr. W.H. Melody
Faculty of Technology, Policy and Management, Delft University of Technology
The Role of Transport Cores in Gene Transcription

An Analytical Framework

TRAIL Research School Cell December 1999

Author
Dr. K.C. Fisher
Dr. K. Kato

Senior Editor
A. M. Smith

Faculty of Information, Media and Management, Cranfield University, UK
Contents

Abstract

1 Introduction ......................................................................................... 1

2 Factors affecting goods transport ...................................................... 2
  2.1 Factors underlying the need for goods transport ......................... 2
    2.1.1 First-order factors .............................................................. 2
    2.1.2 Second-order factors .......................................................... 5
  2.2 Factors underlying the demand for transport services .................. 6
    2.2.1 First-order factors .............................................................. 6
    2.2.2 Second-order factors .......................................................... 8

3 Possible effects of an increase in transport costs ............................... 9
  3.1 Carriers ...................................................................................... 9
  3.2 Shippers .................................................................................... 10

4 Conclusions ..................................................................................... 15

References .......................................................................................... 16

Appendix 1: developments and trends in goods transport ..................... 19
Abstract

It is expected that transport costs will increase considerably in the future. An important indication is the desire of more and more governments to influence the demand for transport through pricing. For instance, many governments consider the introduction of congestion charges on motorways and the EU has plans to charge transport for all the social marginal costs it causes. However, up to now, relatively little is known about how transport demand – and in particular the demand for goods transport - may respond to a substantial increase in transport costs.

This paper explores all the possible ways in which a considerable rise in transport costs may affect goods transport. Currently, shippers in goods transport are generally very price inelastic, because the activities served by goods transport commonly yield benefits that by far exceed transport costs. As a result, the demand for goods transport will probably not respond to small increases in transport costs. But what will happen if governments decide to raise transport costs considerably, as can be expected in the future? Will carriers be able to change their operations in order to offset (part of) the increase in transport costs, e.g. by the use of larger vehicles and the use of ICT? And what will be the effects of a substantial increase in transport costs on the demand for goods transport? Can shippers absorb the higher costs, or will they reduce their transport needs, e.g. by changing their distribution structures in order to decrease average trip length, by reducing the number of shipments by allowing higher inventories or by replacing physical transport by ICT? In this paper, we explore all possible reactions of carriers and shippers to a considerable increase in transport costs. The results are depicted in an analytical framework.
Aprende

The nature and purpose of the development of nuclear energy in various countries and regions are discussed in the context of the current global nuclear energy landscape. The role of international cooperation and agreements in promoting peaceful uses of nuclear energy is emphasized. The challenges and opportunities presented by emerging technologies in the nuclear sector are also highlighted.

In many countries, nuclear energy has played a significant role in the energy mix, contributing to the stability of electricity supply and the reduction of greenhouse gas emissions. However, the development of nuclear energy also raises concerns regarding safety, waste management, and proliferation of nuclear materials.

The importance of nuclear safety is highlighted, with emphasis on the need for robust regulatory frameworks and international standards to ensure the safety of nuclear facilities and the protection of the public. The advancements in nuclear waste management technologies and the development of innovative solutions for the safe disposal of nuclear waste are discussed.

International cooperation through forums such as the Nuclear Energy Agency (NEA) and the International Atomic Energy Agency (IAEA) is crucial in addressing these challenges and promoting the sustainable development of nuclear energy. The role of these organizations in facilitating knowledge exchange, regulatory cooperation, and the sharing of best practices is underscored.

The future of nuclear energy is shaped by the interplay of technological advancements, regulatory frameworks, and societal acceptability. As countries continue to evaluate the role of nuclear energy in their energy portfolios, the importance of transparent communication and public engagement becomes increasingly crucial.

In conclusion, the development of nuclear energy is a complex issue that requires a multidisciplinary approach. By addressing the technical, regulatory, and social dimensions of nuclear energy, countries can strive to harness its potential benefits while mitigating associated risks and challenges.
1 Introduction

There are several indications of a substantial increase in transport costs in the future:

- Several countries plan to charge transport for all marginal social costs it causes, such as environmental costs, congestion costs, infrastructure costs, and traffic accident victims (e.g. EU, 1998; MINISTERIE VAN VERKEER EN WATERSTAAT, 1999);
- More and more cities restrict the amount of (goods) transport entering their centers (e.g. by 'delivery windows');
- Many governments desire private involvement in the financing and operation of (road) infrastructure. Without doubt, this will imply tolling infrastructure users;
- More countries may decide to restrict the amount of transit traffic (Switzerland has already done it, recently the Dutch Council for Transport has suggested to do the same in the Netherlands);
- Many governments (among which the Dutch) have plans to introduce congestion charges in road transport;
- The liberalization of European railways necessitates ‘fair’ entry charges, which cover at least part of the infrastructure costs;
- Fossil fuels such as oil threaten to be depleted in the future, which may give rise to a rationing of stocks.

Most of these indications stem from transport policies that are directly aimed at influencing mobility through pricing. However, little is known about the effects that can be expected as a result of increasing transport costs, especially for goods transport (GEURS AND VAN WEE, 1997: P. 101; CPB, 1998: P. 28). The few studies undertaken in this field notably focus on short-term changes in traffic volumes, given certain price and time elasticities of demand. The mechanisms, through which changes in transport costs may affect goods transport, are more or less regarded as a ‘black box’. Furthermore, structural or long-term effects of an increase in transport costs such as relocation of business activities are commonly ignored, although they will lead to other transport flows.

In this paper we describe how a substantial increase in transport costs may affect goods transport. We distinguish between two types of effects: (1) the effects on carriers' transport decisions and (2) the effects on shippers' transport decisions. The focus is on the latter effects, since most of the above policy measures are aimed at influencing transport demand. The paper has the following structure. In section 2, we explore what factors underlie goods transport. This knowledge will be used in section 3 to examine how an increase in transport costs may affect goods transport. Section 4 contains several conclusions. Appendix 1 contains a description of the major trends and developments in goods transport.
2 Factors affecting goods transport

Goods transport is based upon the need to move certain products at a certain time from one location to another, either for production, inventory or consumption. For instance, the production of assembled industrial products, like cars and PC’s, generally requires a multiplicity of inputs that must be collected from wide-ranging sources. Furthermore, often extensive market areas are served. Transport is the physical thread connecting the company’s geographically dispersed operations and adds value to the company by the physical movement of goods to the place desired and at the time desired (‘time and place utility’; COYLE ET AL., 1996: P. 318). This example illustrates the derived character of goods transport demand.

An analysis of the demand for goods transport is relatively complex because transport demand is affected by many interdependent factors. The factors mentioned in the literature can be aggregated on different levels. First, distinction can be made between factors underlying the need for goods transport and factors underlying the materialization of these needs into the demand for specific transport services. Second, distinction can be made between factors that directly affect the need for goods transport (‘first-order effects’) and those that indirectly affect the need for goods transport (‘second-order effects’). In this section, these dimensions will be used to describe the factors affecting goods transport. Section 2.1 describes the factors underlying the need for goods transport. Section 2.2 describes the factors underlying the demand for specific transport services.

2.1 Factors underlying the need for goods transport

2.1.1 First-order factors

The need to transport goods between certain locations results from the following factors:

1. Demand for final goods: the quantity and quality of goods demanded at a certain time;
2. Geographical dispersal of supply and demand: locations of vendors of raw materials, manufacturers, ancillary suppliers, warehouses, wholesalers, retailers and final consumers;

3. **Horizontal and vertical organization:** product assortment of the company, the number of activities in the supply chain ("links") that are concentrated within the company²,³.

4. **Production structures:** production to stock versus production to order, number of working processes per product;

5. **Logistical concepts:** number of warehouses, inventory sizes, modal choice et cetera.

The Ford Fiesta production network, depicted in Figure 1, illustrates how organization and locations of production facilities lead to transport flows.

**Figure 1: The Ford Fiesta production network in Europe.**

---

² A supply chain is the sequential flow of logistical, conversion and service activities from vendors to final consumers necessary to produce a product or service efficiently and effectively (COOLE ET AL., 1996: P. 9).

³ Companies can choose whether to perform activities themselves or to outsource them. This also applies to transport services. Some companies provide their own transport ("own account") and others outsource their transport to professional transport companies (carriers, forwarders, logistical service providers).
The demand for final goods is generally exogenous to companies, which means that companies have little influence on demand. The other four factors are more endogenous to companies and can be changed according to own insights and needs. These factors are narrowly interrelated, which can be illustrated by the 'just-in-time' concept (JIT), which has become very popular in the last decade. The JIT concept is underpinned by four major elements:

1. Zero inventories (in order to minimize the cost of inventories and the risk of obsolescence);
2. Short lead times (zero inventories imply production to order. Inventories must be available when a firm needs them, not any earlier, nor any later);
3. Small and frequent replenishment quantities;
4. High quality (zero defects).

The JIT concept leads to large volumes of relatively small shipments of raw materials, intermediate goods and final goods. Fast and reliable transport is an essential condition for JIT (COYLE ET AL., 1996: PP. 87-88). In the JIT approach, transport is derived from the production structure and the logistical concept chosen. However, if the transport system worsens (e.g. because transport costs rise or reliability worsens), companies may be forced to reconsider their geographical and organizational structure (new locations, vertical integration) and/or to change their logistical concept (higher inventories).

The factors underlying the need for goods transport have different degrees of flexibility. Locational decisions are generally taken for the long term, since location is coupled with several sunk costs (e.g. transaction costs, employers that are relocated near the company et cetera). Organization, production structures and logistical concepts can be changed more easily, dependent on the amount of fixed capital in use.
2.1.2 Second-order factors

The factors underlying the need for goods transport can in their turn be affected by other factors:

1. **Consumer demand**: consumers' preferences, prices of products and services, demographics, average household income, distribution of income among households (LIPSEY ET AL., 1987: P. 59);

2. **Locations**: locational choice generally depends on a variety of factors. The most well known are: distance to labor-, input- and selling markets, accessibility (or transport costs), costs of land and immovables, agglomeration effects (caused by the proximity of other companies), living climate et cetera (see for instance PELLENBARG, 1999);

3. **Organization**: utilization factors of machinery, economies of scale and scope, situation on ancillary supply markets (market power), uncertainty, transaction costs (including transport costs), coordination costs, the need to be flexible;

4. **Production**: market situation, uncertainty, inventories costs (warehouse costs, risk of obsolescence, risk of lost sales, interest costs et cetera), interest rates, transport costs, product characteristics (life cycle, seasonal influences, risk of obsolescence, value density, perishability et cetera);

5. **Logistics**: transport costs, supply of transport services (quantity and quality), inventories costs, other logistics costs (e.g. coordination, administrative costs), interest rates, product characteristics, organization, geographical dispersal of operations, production structure.

Furthermore, there are factors that apply to all:

1. **Legislation, regulation** and policy (local, regional, national and international): e.g. environmental and safety legislation, labor laws, trade agreements, liberalization of markets, taxes and subsidies;

2. **Investments undertaken in the past** (‘path dependency’);

3. **Technological advances** (e.g. new information and communication technologies and the development of larger and lighter trucks).

---

4 Including the desire to save money for future consumption. Savings are among other things influenced by interest rates.

5 Companies are not always free to choose their locations. For instance, in some sectors as the automotive industry, manufacturers demand that their ancillary suppliers be located nearby to maximize reliability of supply.

6 Including the extent to which legislation and regulations are observed.
2.2 Factors underlying the demand for transport services

2.2.1 First-order factors

The need for goods transport will be translated in the demand for specific transport services, necessary to move a shipment by means at a certain time to a certain location. In the literature, several factors affecting the demand for goods transport services are mentioned. These factors can be summarized as product characteristics, quality demands and transport costs.

Product characteristics

Characteristics of raw materials and intermediate and final products that are generally mentioned in relation to transport demand are (Van Goor et al., 1989, in: De Wit and Van Gent, 1996: pp. 231-232):

- **Physical appearance**: goods can be solid, fluid or gaseous. Furthermore, they can be transported as bulk or in 'unitized' form (pallets, containers, swap bodies);
- **Value density**: the value of the good in the smallest shipment unit possible;
- **Packaging density**: the number of packaged units per volume unit;
- **Volume-weight ratio**: of the smallest shipment unit (cubic meters per kilogram);
- **Perishability**: the period in which the physical or commercial value of the goods remains constant.

Goods with a high value density have relative high opportunity costs (interest cost) and therefore often require fast transport. Furthermore, these goods have high demands on packaging and security in order to minimize the risk of damage.

Quality

The quality of transport services consists of the following elements:

- **Frequency and size of shipments**: these factors are often interrelated. Given a certain production level, a rise in frequency will lead to smaller average shipments;
- **Reliability (or punctuality)**: consistency of the transit time a carrier provides. Not respecting times agreed upon can result in queuing situations at shippers and shortage situations at recipients;

---

7 See footnote 1.
8 Packaging is considered a product characteristic because of its strong dependence on other product characteristics as physical appearance and value.
9 In a later publication, Van Goor leaves out the commercial value in his definition of perishability, although it is unclear why (Van Goor et al., 1996: pp. 53-54).
10 Note that the market value has significant influence on the transport options. High-valued goods can permit to be transported by more expensive modes than lower-valued goods and are therefore more flexible (Stefansson and Tilanus, 1998: p. 92).
- **Transport time:** the total time between the moment the shipper makes the goods available and the receipt of the shipment. Transport time is a direct result of distance, travel speed and organization of transport (e.g. direct transport versus hub-and-spoke transport). Transport time directly influences the size and cost of the inventory in transit as well as at the recipients' safety inventory;

- **Security:** the extent to which goods arrive in the same condition as they were when delivered at the carrier;

- **Supportive services:** up-to-date information on the transport (e.g. through EDI and tracking-and-tracing systems), administrative services (customs) et cetera;

- **Flexibility:** the carrier's ability to adapt to unforeseen circumstances or changes in shippers' demands (e.g. by disposal of multiple transport modalities or networks).

### Transport costs

Transport costs differ according to whose point of view is considered (STEFANSSON AND TILANUS, 1998: P. 4). For carriers and companies operating 'own account' transport services, transport costs are the costs they bear to transfer the goods from A to B under the shippers' conditions. For shippers, transport costs consist of the price they pay to the carriers (or forwarders if these are involved) and other, indirect costs as inventory cost of the goods in transit, cost of damage et cetera.

In this paper, we focus on the transport costs as perceived by shippers, because shippers are the demanders of goods transport services. Distinction is made between **direct** and **indirect** transport costs:

1. **Direct transport costs:** since most shippers have outsourced transport, their direct transport costs consist of the price they have to pay to carriers or forwarders for moving goods from a certain location to another under certain conditions. The price is generally related to the underlying operating costs (fuel, labor, maintenance, overhead, writing-off and interest, profit);

2. **Indirect transport costs:** all other costs, borne by the shipper, and related to the transport of goods. These costs consist of opportunity costs and damage. Opportunity costs arise when resources are used in a particular venture and consist of the benefit forgone by not using the resources in their best alternative use (LIPSEY ET AL., 1987: P. 165). In transport, opportunity costs are commonly the capital cost of the goods in transit (interest) and the capital costs of vehicles and other transport equipment. **Damage** is goods' value loss due to physical damage or devaluation of commercial value of goods in transit, and 'scheduling costs', the cost that arise when goods are not delivered in time (lost sales, forgone productivity due to delays in production).

Direct and indirect transport costs are indirectly related. Prices of transport are generally higher when more security is demanded (a quality aspect), but damage costs are generally lower.

TRAIL Research School, December 1999
Interrelations between factors underlying the demand for transport services

Product characteristics, quality demands and transport costs are interrelated. The product characteristics in principle determine the quality of the transport service demanded. If carriers can offer that quality, the price a shipper is willing to pay for the transport service depends on the ratio between total transport costs (price included) and the time and space utility derived from the service (the benefits). If the costs are lower than or equal to the benefits, the transport service will normally be bought; otherwise, the quality demands and even the transport needs can be adapted.

2.2.2 Second-order factors

Product characteristics can change as a result of a changing demand for final goods or technological advances.

The quality of transport services can be affected by the following second-order factors:

- **Frequency and size of shipments**: changes in production level, size of desired inventories at shippers and recipients, transport costs;
- **Reliability**: congestion and other (unforeseen) delays, quality of vehicles and infrastructure;
- **Transport time**: ditto;
- **Security**: packaging, care of handling, quality of vehicles used, safety of driving;
- **Flexibility**: number of transport modalities and networks the carrier has disposal of, legislation on working and resting hours, reserve capacity (personnel, vehicles).

Finally, the following factors can affect transport costs:

1. **Direct transport costs**: input prices (fuel et cetera), taxes and subsidies, infrastructure charges and other charges, market situation (competition, market power, price discrimination), congestion and other delays, possibilities of back-hauling;

2. **Indirect transport costs**: opportunity costs are affected by interest rates and market situation (profit opportunities) and travel time (dependent on traffic). Factors affecting damage are congestion and other delays (affected by traffic volumes), product life cycles and the risk of lost sales (affected by the demand for final goods), quality of transport services (security, reliability).

Further, there are several second-order factors that apply to all (see section 2.1.2).

---

11 These benefits can be expressed as the difference between the products' selling price and their production costs.
3 Possible effects of an increase in transport costs

In section 1 we observed several indications of a substantial increase in transport costs in the future. In this section we examine how such an increase in transport costs may affect goods transport flows. Following Stefansson and Tilanus, we expect that an increase in transport costs will work through two stages (STEFANSSON AND TILANUS, 1998: P. 5). First, a rise in transport costs will affect carriers' costs and may lead to changes in carriers' transport decisions. For instance, the carrier may decide to use other vehicles, other modalities or other routes in order to offset (part of) the rise in transport costs. However, in case of a substantial increase in transport costs, it is likely that carriers have to charge part of the increase in transport costs to the shipper. Then in the second stage, higher carrier prices may lead to a reconsideration of shippers' demand for transport services. For instance, shippers may decide to allow higher inventories, which leads to less goods transport.

This section is structured as follows. In section 3.2 the possible effects of an increase on carriers' transport decisions are described. In section 3.3 the possible effects for shippers' transport decisions are treated. Figure 3 presents an analytical model that depicts all the possible reactions of carriers and shippers to a rise in transport costs.

3.1 Carriers

An increase in transport costs will first affect the carriers' costs. In principle, carriers have three options: simply pass on the extra costs to their customers, bear the costs themselves or try to absorb the extra costs by avoiding them or by raising efficiency. In competitive markets the former solution is usually not possible, since carriers would price themselves out of the market. Carriers then will have to absorb the extra costs and pass the rest on to shippers. They can absorb part of the rise in costs by: 12,13

1. Changes in vehicle size (if costs vary less than proportionally with truck size);
2. Modal shift (substitution to a cheaper modality);
3. Raise efficiency (higher load factors);

12 It is assumed that in case of own account transport, the same reactions will occur, although within the shippers' organization.
13 For instance, the Dutch road haulage market is very competitive and carriers have little market power vis-à-vis shippers. It might be possible that shippers will not accept any increase in prices so that these carriers will have to bear the costs themselves. New ways of logistical cooperation between carriers and shippers – like 'open book accounting' - favor this expectation.
14 Dependent on the way the extra charges are collected, carriers can also avoid the extra costs on illegal ways.
4. Shift in time (later/earlier);
5. Route choice;
6. Raise 'environmental efficiency' (cleaner motors).

However, given a substantial increase in transport costs it is likely that carriers will have to pass on part of the cost increase to the shippers.

### 3.2 Shippers

In general, the share of direct transport costs in total costs, turnover and total value added is very low (see Table 1). This phenomenon raises the expectation that industries generally will be fairly insensitive to increases in transport services’ prices. However, profit on turnover may be small, in which case an increase in transport costs could have a significant impact on the profit margin (Stefansson and Tilanus, 1998: P. 44; Bleijenberg, 1998: P. 27). Furthermore, in competitive markets, companies normally can not pass on an increase in transport costs to the customers. Shippers then will have to try to offset the increase in transport costs by reconsidering their demand for transport services.

**Table 1: Logistical costs as a percentage of turnover (Europe)**

<table>
<thead>
<tr>
<th>Product</th>
<th>Storage (2)</th>
<th>Inventories (3)</th>
<th>Transport (4)</th>
<th>Administration (5)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food, beverages, tobacco:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- producer</td>
<td>2.2</td>
<td>2.8</td>
<td>3.7</td>
<td>1.7</td>
<td>10.4</td>
</tr>
<tr>
<td>- wholesaler</td>
<td>3.5</td>
<td>2.4</td>
<td>3.2</td>
<td>1.8</td>
<td>10.9</td>
</tr>
<tr>
<td>- retailer</td>
<td>3.5</td>
<td>1.4</td>
<td>2.4</td>
<td>2.0</td>
<td>9.3</td>
</tr>
<tr>
<td>Chemical industry:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- producer</td>
<td>2.3</td>
<td>2.6</td>
<td>3.8</td>
<td>1.5</td>
<td>10.2</td>
</tr>
<tr>
<td>- wholesaler</td>
<td>2.7</td>
<td>2.8</td>
<td>3.0</td>
<td>2.0</td>
<td>10.5</td>
</tr>
<tr>
<td>- retailer</td>
<td>2.2</td>
<td>3.8</td>
<td>2.1</td>
<td>1.2</td>
<td>9.3</td>
</tr>
<tr>
<td>Computers/electronics:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- producer</td>
<td>2.0</td>
<td>3.8</td>
<td>2.0</td>
<td>2.5</td>
<td>10.3</td>
</tr>
<tr>
<td>- wholesaler</td>
<td>3.8</td>
<td>3.6</td>
<td>2.5</td>
<td>2.3</td>
<td>12.2</td>
</tr>
<tr>
<td>- retailer</td>
<td>3.3</td>
<td>2.5</td>
<td>1.6</td>
<td>2.0</td>
<td>9.4</td>
</tr>
<tr>
<td>Paper and paper ware:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- producer</td>
<td>3.0</td>
<td>3.6</td>
<td>4.7</td>
<td>2.1</td>
<td>13.4</td>
</tr>
<tr>
<td>- wholesaler</td>
<td>3.2</td>
<td>2.5</td>
<td>2.5</td>
<td>1.0</td>
<td>9.2</td>
</tr>
<tr>
<td>- retailer</td>
<td>2.9</td>
<td>1.3</td>
<td>1.5</td>
<td>0.8</td>
<td>6.5</td>
</tr>
<tr>
<td>Average - producer</td>
<td>2.3</td>
<td>2.9</td>
<td>3.0</td>
<td>1.9</td>
<td>10.1</td>
</tr>
<tr>
<td>Average - wholesaler</td>
<td>3.0</td>
<td>2.9</td>
<td>2.9</td>
<td>2.2</td>
<td>11.0</td>
</tr>
<tr>
<td>Average - retailer (5)</td>
<td>3.0</td>
<td>2.0</td>
<td>2.3</td>
<td>1.6</td>
<td>8.9</td>
</tr>
</tbody>
</table>

Options differ for general versus specific rises in transport costs. Specific rises (e.g. congestion charges during peak hours) are easier to avoid than general rises in transport costs (CPB, 1997: PP. 312-315).

However, the contribution of transport costs to total costs varies along products. Furthermore, transport costs are relatively low for manufactured goods, but relatively high for raw materials and inputs like timber (Jansson and Lindberg, 1997: P. 61).
Notes: (1) percentages of logistical costs of producers, wholesalers and retailers cannot be added, because these parties may not be equally present; (2) storage costs are the costs that are made to store goods (warehouse, handling, internal transport et cetera); (3) costs of inventories consist of interest costs of the goods in inventory and the costs of goods losing their commercial value; (4) transport costs are prices paid to carriers and insurance costs (direct transport costs – see section 2.2.1); (5) administration costs are costs for supporting the physical distribution (administration, system costs and order processing); (6) averages including four other sectors (electronic machinery, car industry, engines/metal, pharmaceuticals). Source: VAN DEN BOSSCHE ET AL., 1995: APPENDIX 2).

However, an increase in transport costs may also have other effects for shippers, since changes in transport costs may affect other transport, such as passenger transport, which may result in changes in circulation, travel time and reliability. This may in turn affect shippers' direct transport costs (if carriers' prices depend on transit time) and his indirect transport costs (notably scheduling costs; see section 2.2.1).

Figure 2 gives an indication of the reactions of different categories of traffic on changes in transport prices, travel times and reliability. Business travelers are generally insensitive to changes in transport prices, but are relatively sensitive to changes in travel times and changes in arrival times. The price-insensitivity can be explained by the fact that generally to an employer, labor costs are higher than transport costs. This also explains the sensitivity for changes in arrival times; arriving too late may lead to productivity loss (scheduling costs). The opposite holds for travelers with recreational travel motives. An increase in travel costs will therefore generally lead to a decrease of the latter category of transport. This is confirmed by several studies on the effects of increases in transport costs, for instance due to the introduction of congestion charges (see for instance VAN DER VLIET ET AL., 1998).

**Figure 2:** Sensitivity in road transport to changes in transport costs and travel times, differentiated to traffic categories.

<table>
<thead>
<tr>
<th>Transport motive</th>
<th>Direct transport costs</th>
<th>Travel time</th>
<th>Reliability of arrival/departure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passenger - business</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Passenger - commuters</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Passenger - other</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Goods transport- general</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

Source: CPB, 1998: P. 20 (passenger transport); authors (goods transport).
The way in which shippers will respond to an increase in transport costs will therefore probably depend on how both direct and indirect transport costs are affected. Given that direct and indirect transport costs will change significantly, it is likely that trade-offs between transport costs and the benefits of activities that that are served by goods transport will change. The most well known trade-offs are (COYLE ET AL., 1996; HEPWORTH AND DUCATEL, 1992; JACOBS, 1998; RUIJGROK, 1991; US DoT, 1996)\(^7\):

1. **Locational choice**: transport costs versus relocation benefits (agglomeration effects, better living climate et cetera);

2. **Production**: transport costs versus central production (economies of scale) and increased flexibility. Central production may yield economies of scale that outweigh the higher transport costs compared to decentral production. Production to order yields higher transport costs than production to stock (more shipments, a higher speed is required which is more expensive, possible loss of economies of size by the use of large vehicles), but leads to lower interest costs and a reduced risk of obsolescence;

3. **Organization**: transport costs versus the benefits of in- or outsourcing. If a company already has outsourced several activities, a trade-off can be made between higher transport costs versus lower costs of materials sold by suppliers located at larger distance. The trade-off between own account and professional transport depends on a comparison of total transport costs;

4. **Logistics**: direct transport costs versus inventory costs. A decrease of inventory size at shippers, in transit and at recipients normally leads to a higher frequency of smaller shipments and thus to higher transport costs (more transport and possible loss of economies of size by the use of large vehicles). Further, the number of warehouses influences transport costs. Central warehousing leads to lower inventory costs but higher transport costs (larger average distances, loss of economies of scale through larger vehicles et cetera);

5. **Quality**: transport costs versus frequency and size of shipment, reliability, transit time (faster modalities are commonly more expensive), security, supportive services and/or flexibility. A higher quality may lead to higher direct transport costs but lower indirect transport costs;

\(^7\) Usually, decisions involve more than one trade-off. For instance, if shippers choose for a certain transport modality, they compare direct transport costs to indirect transport costs (consequences for inventory costs, risk of damage et cetera; COYLE ET AL., 1996: P. 318).
6. **Transport services**: transport costs versus costs of substitutes (e.g. information and telecommunication technology). Currently, cut flowers are send by airplane to the auction in Aalsmeer where they are sold and flown to their new destinations. If transport costs would increase (e.g. if VAT and excise-duty are levied on kerosene), then it may become cheaper to switch to an electronic auction and fly the flowers directly to the buyer.\(^\text{18}\)

Given a substantial increase in transport costs, it is likely that shippers will reconsider some of the above trade-offs and with that, their demand for transport services.

We have now treated all the possible reactions of carriers and shippers to a substantial increase in transport costs. Figure 3 summarizes all these reactions. The Figure can be used as an analytical framework to predict the effects of a substantial increase in transport costs on certain sectors.

\(^{18}\) In practice, substitution of goods transport by telecommunication seems to be possible only to a limited extent. Telecommunication is notably an alternative for passenger transport (e.g. 'teleworking' and 'teleshopping'; HEWORTH AND DUCATEL, 1992: P. 29).
Figure 3: Analytical framework: possible effects of a substantial increase in transport costs.

Demand for final goods
- Location
- Quality
- Time
- Price (maximum)

Supply of final goods

Production
- Location
- Organization
- Prod. structure

Inventories
- Location
- Size/velocity

Alternatives to transport
- Quality
- Price

Indirect transport costs

Transport needs
- Quality
- Time
- Route
- Etc.

Direct transport costs

Transport services
- Quality

Carriers' costs

Purchase of transport services
- Quality
- Etc.
- Modal choice

Purchase of alternative services

Infrastructure
- Quality
- Capacity

Goods traffic flows

Other traffic

Circulation

Rise in transport costs

Transport prices

The Role of Transport Costs in Goods Transport
4 Conclusions

There are several indications that transport costs will increase considerably in the future. An important indication is that more and more governments plan to raise transport costs in order to influence transport demand. However, still little is known about how transport demand – and in particular the demand for goods transport – will respond to a substantial increase in transport costs.

In this paper, we examined the ways in which a substantial rise in transport costs may affect goods transport. First, we explored the drivers behind goods transport. We distinguished between factors that underlie the need for goods transport and factors that underlie the demand for specific transport services. Examples of the former factors are geographical dispersal of supply and demand and logistical concepts, examples of the latter factors are transport costs and quality demands.

Second, we explored how a substantial increase in transport costs may affect the need for goods transport and the demand for goods transport services. We assumed that an increase in transport costs will work through two stages:

1. A rise in transport costs will first affect carriers’ costs and may lead to changes in carriers’ transport decisions. However, it is likely that carriers have to charge part of the increase in transport costs to the shipper;

2. A substantial increase in transport costs will affect shippers’ direct and indirect transport costs, which may lead to changes in shippers’ transport needs and demand for transport services.

Finally, all the possible reactions of carriers and shippers to a substantial rise of transport costs were summarized in an analytical framework.
References


- Geurs, K.T. and G.P. Van Wee (1997). Effecten van prijsbeleid op verkeer en vervoer (Effects of pricing on transport and traffic; in Dutch), Bilthoven: RIVM.


• NEI/NEA (1990). Vervoerwijzekeuze in het goederenvervoer (Modal choice in goods transport; in Dutch), Rotterdam/Rijswijk: NEI/NEA.


• Stefansson, G. and B. Tilanus (Eds.) (1998). State of the art and freight cost database, SOFTICE project (Survey on freight transport including cost comparison for Europe), Brussels: European Commission.


• Wit, J. de and H. van Gent (1996). Economie en transport (Economics and transport; in Dutch), Utrecht: Lemma B.V.
Appendix 1: Developments and trends in goods transport

Introduction

In section 2, the factors underlying the demand for goods transport were treated. This section deals with the dynamics in these factors. In view of the amount of factors, it is a major task to treat all the developments and trends. Furthermore, it is expected that the dynamics will vary between companies, sectors and even regions. Therefore, only the most important developments and trends, as mentioned in the literature19, are treated. The focus is on Western Europe. First, the major developments and trends related to the demand for goods transport are treated. Then, the consequences for production and logistics are described.

Major development and trends

Four interrelated elements are generally regarded as the primary drivers of economic developments in the last decades: deregulation20, new consumer preferences, technological advances and an increased care for the environment. Due to deregulation production and supply have become more internationally oriented ('globalization' and 'regionalization'). The resulting increase in competition has commonly reduced prices and has led to a shift of market power from producers to consumers. New consumer preferences have made demand more divers and therefore more difficult to predict. Consumers' tastes change quickly and therefore demand has become relatively capricious. As a result, product life cycles become shorter. For instance, in consumer electronics, life cycles of 6-9 months become more and more usual21. Therefore, companies are constantly forced to innovate their products. Further, there is a trend toward lighter and smaller products with high value density. Technological developments have led to improvements in efficiency of production, logistics and transport and with that to a shift of industrial activities to services ('dematerialization of production', 'new economy'). Lastly, more severe environmental legislation has forced industry to produce less polluting and to reuse materials (recycling).

Consequences for production

The increased competition and the shortening of product life cycles have forced many companies to make production more 'flexible'. Commonly, four strategies are used.

20 Deregulation means that market barriers are reduced, causing markets to become more open to competing companies and/or allowing companies to develop other economic activities.
21 A result of the changes in demand is that within supply chains, the power of retailers vis-à-vis producers has increased because they have direct contact with consumers and therefore have knowledge of demand (VAN GOOR ET AL., 1996: P. 10).
First, inventories are minimized in order to minimize the risk of obsolete goods and to decrease inventory costs. Therefore, many production companies switched from production to stock, aimed at achieving economies of scale, to production to order. Lead times are minimized to offer high customer service. Second, many products are designed on a modular base, using many standard parts in order to be able to adapt products to customers' needs ('mass customization'). Third, companies are constantly trying to improve efficiency, e.g. by using modern information- and communication technologies, by outsourcing activities that do not belong to their core activities (and that can be executed cheaper by specialized companies) and by choosing the optimal location for their operations. Finally, companies within a supply chain to an increasing extent cooperate in order to respond quickly on a changing demand, improve the customer service and to lower total costs ('supply chain management', 'co-makership').

Consequences for logistics

The changes in production have important consequences for logistics. Production to stock, the increased outsourcing of activities, the strive after short lead times and the increased cooperation within supply chains are only possible when products and inputs are available when they are needed. Therefore, just-in-time concepts and flexible transport systems are increasingly used. Second, inventories have been minimized, leading to a centralization of distribution (less distribution centers). Third, the increased recycling of inputs (e.g. packaging) and goods has led to other logistical concepts ('reverse logistics'). As a result of all these developments, transport and traffic flows increase. The major part of these flows is carried out by road haulage, which however causes severe environmental problems. Fourth, the increasing outsourcing of logistical activities by large companies has led to the rise of logistical operators, who take complete responsibility for all the distribution activities of their customers.

22 Customer service (intangible qualities of a product, as flexibility in delivery, information on order status, availability of spare parts etc) has become an increasingly important competition factor (De Wit and Van Gent, 1996: pp. 370-371).
23 Further, by adding value in a late stage of production, the interest costs of inventories are reduced.
24 This trend is particularly noticeable in the European car industry, where the large car manufacturers (e.g. Ford) tend to concentrate on the design of new cars, marketing and sales and outsource manufacturing and assembling activities (Nieuwsblad Transport, August 10, 1999).
25 In Europe big changes in distribution structures have appeared after 1992, the opening of the internal market (Adviesdienst Verkeer en Vervoer, 1997: p.50). The reason is probably that economies of scale can be realized by combining distribution centers (Ruijgrok, 1991: p. 14).
26 However, this does not mean that all transport needs can be fulfilled. Distribution within cities in The Netherlands has become problematic, since distributors are confronted with tighter 'delivery windows' and increasing congestion. In some cases, this has led to a restructuring of distribution. Suppliers of large retailers bring their shipments to a central distribution center, where the shipments are combined and transported to the retailer (Adviesdienst Verkeer en Vervoer, 1997: p.51).
Further, in logistics there is an increased use of information and communication technologies (e.g. EDI and Internet). This may lead to a higher transport efficiency (i.e. less transport due to bundling of shipments, shorter routes, reduction of empty hauls) and may favor concepts as the 'direct distribution' from producer to individual consumer of 'individual order-based products'. Lastly, also European transport markets are affected by the trend of deregulation and privatization. It is expected that this will result in improved quality of transport services and lower prices.
Consequences for logistics

The changes in production have important consequences for logistics. Production in
stock, the increased outsourcing of activities, the drive after lowest costs and the
increased cooperation within supply chains are only possible when products and
materials are delivered when they are needed. Therefore, just-in-time concepts and
flexible transport systems are increasingly used. Second, shortening of product cycles
has led to a fragmentation of distribution (less distribution categories) and
increased recycling of inputs (e.g., packaging) and goods that are returned to
location logistics. As a result of all these developments, transport and
logistic flows have increased. The major part of these flows is carried out by road
transport which however causes severe environmental problems. Fourth, the increasing
outsourcing of logistical activities by large companies has led to the role of logistic
operators who take complete responsibility for all the distribution activities of their
customers.

---

* Customer service (intangible qualities of a product, service delivery, information on stock availability, delivery date, etc.) has become increasingly important in competition. (Euker, 1994; Womack, 1996, Kim, 2001, 2007)

* However, as the price of logistics is rising, the demand for lower costs is escalating. (Hann, 1998; Hennigan, 1997; Thompson, 1998)

* The trend is perfectly visible in the European car industry, where the large car manufacturers

  * Nokia: in cooperation on the design of new cars, mastering and using such new technologies in manufacturing and assembling vehicles (Hann, 1998; Thompson, 1997, 1998)
Located in Jozietsr and Scipio Aipbo
A first step towards a conceptual framework

TRAIL Research School, Dell December 1999

Author

Dean, P.M. Maling

Co-Sponsoring: Department of Neural Communication, Energy Transformation, Environment, Information

Field of Education: Business Administration, Economics, Environmental Economics, Science, Technology, Agriculture, Environment

Editor of Safety: Economics

Editor of Agriculture: Economics

Editor of Information: Economics
Abstract

Contents

1. INTRODUCTION

2. ABOUT GROWTH EXPECTATIONS FOR SCHIPHOL AIRPORT
   2.1 Introduction ...................................................... 2
   2.2 World air traffic volume ........................................ 2
   2.3 Schiphol Airport ............................................... 3

3. SUB-OPTIMAL LOGISTICS IN THE SCHIPHOL AREA ......... 5

4. LOCKED-IN LOGISTICS
   4.1 Introduction .................................................... 7
   4.2 Locked-in logistics further defined .......................... 7

5. PREVENTING AND OVERCOMING LOCKED-IN LOGISTICS .... 9
   5.1 Introduction ..................................................... 9
   5.2 Spatial behavior of companies ................................ 9
   5.3 Logistics behavior of companies ............................. 10
   5.4 The decision making process and locked-in logistics .... 12

6. CONCLUSIONS AND FURTHER RESEARCH ..................... 14

REFERENCES ......................................................... 15
Abstract

If we focus on the Schiphol Airport area, and within that area on freight transport, modes of transportation and company locations, it seems that major logistics sub-optimizations occur or are going to occur. It seems that companies are situated at "wrong" locations and that "wrong" modes of transportation are used. These logistics sub-optimizations are not easy to solve. They seem to be fixed. It seems that some logistics sub-optimizations are "locked-in" in our transport system. Locked-in logistics deals with this kind of logistics sub-optimizations. The question is: How to overcome and prevent them? This paper is a first step towards a framework for analyzing, overcoming and preventing locked-in logistics situations within the Schiphol area.
World air traffic volume is growing rapidly. The Dutch situation shows the same. The Dutch aviation is growing rapidly and Schiphol Airport is an attractive location factor for all kinds of companies (see paragraph 2). At the same time, society puts environmental limits for air transport, Schiphol and the Schiphol area. Until 2010, Schiphol Airport will have to function at the Schiphol location within those environmental borders. At this moment, Schiphol and the Schiphol area are already experiencing problems in the form of lack of space, tight environmental constraints, and congestion. Therefore, Schiphol Airport and the Schiphol area have to follow a selective growth strategy. This means, among other things, avoiding sub-optimal logistics situations.

Logistics sub-optimizations can lead to bad logistics services offered by the Schiphol area. If we focus on the Schiphol area, and within that area on, freight transport, modes of transportation and company locations, it seems that major logistics sub-optimizations occur or are going to occur. It seems that companies are situated at "wrong" locations and that "wrong" modes of transportation are used (see paragraph 3). These logistics sub-optimizations are not easy to solve. They seem to be fixed. It seems that some logistics sub-optimizations are "locked-in" in our transport system. Locked-in logistics deals with this kind of logistics sub-optimizations (see paragraph 4). The question is; How to prevent them and how to overcome them? This paper offers a first step towards a framework for analyzing, preventing and overcoming locked-in logistics.

With respect to locked-in logistics situations in the Schiphol area, we follow two lines of reasoning (see paragraph 5): a) preventing future logistics sub-optimizations and locked-in logistics situations by changing the spatial behavior of companies (management of company locations) and, b) to reduce or overcome current locked-in logistics situations by changing the logistics behavior of companies (management of modes of transportation).

We can study the spatial and logistics behavior of companies from the perspective of the physical infrastructure (buildings, roads, airports, cars, etc.). We also can study the spatial and logistics behavior of companies from the perspective of the social infrastructure (institutions, social networks, structures of knowledge, rituals, paradigms, etc.). Both infrastructures are connected. We will aim at influencing the physical infrastructure via the social infrastructure. The decision making process is part of the social infrastructure. We will explore three theoretical approaches from the field of decision making to study locked-in logistics (see paragraph 5.4).
2. About growth expectations for Schiphol Airport

2.1 Introduction

First, in paragraph 2.2, the growth of the world air traffic volume will be discussed. This shows a growth trend in air transport and some consequences from that. The growth of the world air traffic volume has a big impact on the growth of Schiphol Airport. The growth of Schiphol Airport and its logistics problems are discussed in paragraph 2.3.

2.2 World air traffic volume

In his book "The rise and fall of infrastructures", Grübler studied the growth and diffusion patterns of transport systems. For the world air transport he comes to the following conclusions [Grübler; 1990, p 162-171].

Conclusion 1:
The world air transport volume, which includes cargo, air-mail and passengers, can be expressed in ton-kilometers (t-km) transported annually. The growth of t-km transported by all air carriers increased at an exponential rate from 1950 up to 1980. The estimated saturation level of around 290 billion ($10^9$) t-km will be reached after 2010. Further market volume growth which appears likely after the year 2000 will bring a possible second expansion phase. This growth trend in world air traffic will have important consequences on the necessary support infrastructures as air traffic control, airways, airports and hubs. The growth of air traffic could therefore be confronted with serious bottlenecks as congested air corridors and airports in many areas of the world.

Conclusion 2:
Within the general growth in total market volume in t-km, one can note a gradual shift in the three different types of payload transported by air. If we take 1950 as a start, a gradual decrease in importance of mail and a structural shift between passenger and cargo can be seen. From 1950, the market share (percentage of the world wide t-km transported by aircraft) of air-mail is decreasing, the market share of passengers is decreasing, and the market share of cargo is increasing. The "just-in-time" principle and the continuing request to reduce inventories will increase the importance of transport speeds and flexibility. Air cargo may not be confined to traditional high value or perishable goods but may enter into markets considered until now as unimaginable for air transport.
Conclusion 3:
The comparative advantage of air transport as regards speed, flexibility and quality of service appears to be so high that external events have not significantly affected the market volume expansion. It is worth noting that the regular growth path in world air traffic was not influenced significantly by the rapidly rising fuel costs.

Conclusion 4:
The growth of air transport neither appears to be constrained by external events like rapidly rising fuel costs nor by the responses of competitors in long distance traveling. The only remaining constraint to be analyzed is whether the development of aircraft infrastructure did or could in the future constitute a bottleneck for the growth of air traffic.

2.3 Schiphol Airport

We can find air transport at the beginning of its growth cycle. This corresponds to the growth figures and growth expectations of the Dutch situation. An illustration:

In 1986, Schiphol handled 12 million \(10^6\) passengers and 0.4 million tons air cargo. The number of air transport movements in 1986 was 200 thousand. In 1998, Schiphol handled 34.4 million passengers and 1.2 million tons of air cargo. The number of air transport movements rose to 377 thousand in 1998. [Bouwens; 1996, p 359, 565, 374] [Schiphol Group, 1998]

In December 1998, the Dutch government allowed Schiphol to grow till 60 million passengers, 2.6 million tons of air cargo and 600 thousand air transport movements. These numbers will be reached in the year 2010. After 2010 additional airport capacity will be needed. Plans for an island in the North Sea are a very serious option.

All the scenarios show an enormous growth of air transport for the Dutch situation. The most recent scenario (the EC+ scenario from the BCI/NEI-research from 1997) predict that air transport in The Netherlands in the year 2025 will accommodate 103 million passengers and 7.7 million tons air cargo. [Oosterhaven; 1998, p 101]
Schiphol is a part of an international and national transport network and performs a role of a hub. Schiphol provides connections between many different modes of transport. Schiphol as an inter-modal and multi-functional hub, forms an attractive location factor for all kinds of companies. The two most important Schiphol-studies, who studied the economic effects of Schiphol for the Dutch economy are: The IEE-research of 1993 and the BCI/NEI-research of 1997. Both studies use the following classification for economic activities in the area of Schiphol [Kramer; 1998, p 10]:

- **Direct airport related activities:**
  These activities are performed at Schiphol Airport.

- **Indirect airport related activities:**
  This are the so called backward and forward linkages. Backward linkages are the suppliers of the companies working at Schiphol Airport. An example of forward linkages are European Distribution Centers (EDC) who are making use of air cargo or international offices who are using Schiphol Airport for their business flights.

Some important conclusions about Schiphol Airport related companies are [Kramer; 1998, p 11-14]:

- most of the airport related companies are part of the transport and distribution sector,
- the percentage of service oriented companies is rising fast,
- 75-80% of the airport related companies are concentrated in an area with a radius of 30-60 kilometers around Schiphol (Schiphol area),
- in the Schiphol area the airport related companies are often clustered and located along transport corridors,
- at least 30-40% of the airport related companies is using Schiphol for both, air cargo and business trips.

We can conclude that the Dutch aviation is growing and that Schiphol is an attractive location factor for all kinds of companies. But, at the same time, the society puts environmental limits for air transport, Schiphol and the Schiphol area. Until 2010 Schiphol Airport will have to function at the Schiphol location within the environmental borders. Schiphol Airport and the Schiphol area have to follow a selective growth strategy. At this moment Schiphol and the Schiphol area are already experiencing problems in the form of: lack of space, tight environmental constraints, and congestion (on the roads, on the airport, and in the air) [BCI/NEI; 1997, p 84]. A selective growth strategy means, among other things, avoiding sub-optimal logistics situations.
3. Sub-optimal logistics in the Schiphol area

The systems approach holds the basic belief that integrated system performance can and will produce an end result greater than is possible from non-coordinated performance. All functions or activities need to be understood in terms of how they affect, and are affected by, other elements and activities with which they interact. The idea is that if one looks at actions in isolation, one will not understand the big picture or how such actions affect, or are affected by, other activities. Without considering decisions on the larger system, sub-optimization often occurs. Logistics sub-optimization means that the individual activities in the system are operating well but the net result on the total system is relatively poor performance. [Lambert; 1998, p 9-10] [Magee; 1985, p 31-32]

If we define the Schiphol area as the total system and if we concentrate on freight transport, modes of transportation, and company locations, it seems that major logistics sub-optimizations occur or are going to occur. It seems that companies are situated at "wrong" locations and that "wrong" modes of transportation are used. In other words: the Dutch aviation is growing rapidly and it seems we are choosing wrong about modes of transportation to be used and company locations in the Schiphol area. This can lead to bad logistics services offered by the Schiphol area. Within this framework, four airport products of a mainport are distinguished [BCI/NEI; 1997, p 79]: 1) the destinations and flight frequencies offered by Schiphol Airport, 2) the handling of passengers and air cargo at the airport terminals, 3) the accessibility of Schiphol Airport over land via roads and railways and, 4) the number and international orientation of companies in the neighborhood of Schiphol Airport. Below follow three examples of possible logistics sub-optimizations occurring in the Schiphol area:

Example 1
About 50% of the seaport oriented European Distribution Centers (EDC's) in The Netherlands is located within a radius of 50 kilometres around Schiphol [BCI/NEI; 1997, p 82]. An explanation for this is the importance of the factor time for air cargo, although air cargo is only a small part of the total goods flows for these EDC's. The individual EDC's optimized their own logistics processes. For the Schiphol area as a whole this will lead to logistics sub-optimizations. Scarce space in the Schiphol area is used for seaport oriented companies. It must be noted however that a part of the seaport oriented companies may be oriented towards the seaport of Amsterdam.

Example 2
A lot of goods flows are going from airport Schiphol to seaport Rotterdam and the other way around. In 1994, 1,5 million tons of goods were transported from Rotterdam to Schiphol and 0,3 million tons of goods from Schiphol to Rotterdam. These figures show a degree of distortion because the transport of kerosine is included [BCI/NEI; 1997, p 81].
This example connects with the example above. From the fact that a lot of goods flows are going from Rotterdam to Schiphol and from Schiphol to Rotterdam can be concluded that around Schiphol many seaport oriented companies are located. For the Schiphol area this means a logistics sub-optimization because of extra transport and inefficient use of the scarce capacity of the transport corridors.

Example 3
A company location close by Schiphol Airport can lead to a company strategy of using air cargo as a safety-net for all kinds of logistics problems. Fast physical distribution via air transport becomes the general solution for bad logistics management. This can be an optimal strategy for an individual company. But, for the Schiphol area as a whole, however, this is not the case because the scarce capacity of Schiphol Airport is partly used for air cargo that easily could have been transported by other means of transport. This assumption finds some support of Oosterhaven [Oosterhaven; 1998, p 114-119]. Oosterhaven asked 26 transport and distribution companies, located in the Schiphol area, what the effect for the company would be if, on a short term, the number of intercontinental flights from Schiphol Airport would be halved or doubled. When halving takes place, only 6 of the 26 companies expect some small effects. When doubling takes place, only 3 of the 26 companies expect some small effects. An interesting question is how the companies will deal with halving. By better logistics management?
4. Locked-in logistics

4.1 Introduction

The logistics sub-optimizations, discussed in paragraph 3, bring disadvantages for: companies located in the Schiphol area, companies who want to move to the Schiphol area, the development of Schiphol Airport, the local authorities in the Schiphol area and, the national government. If these sub-optimizations bring disadvantages for all parties, they must be easy to solve. This is not true! The sub-optimizations seem to be fixed.

The policies of the Ministry of Transport, Public Works and Water Management are often directed at preventing or overcoming logistics sub-optimizations as discussed in paragraph 3. For instance, the modal shift policy. Till now the results of the modal shift policy are poor.

It seems that some logistics sub-optimizations are "locked-in" in our transport system. Locked-in logistics deals with this kind of logistics sub-optimizations. How to overcome them and how to prevent them. This paper is specially focused on locked-in logistics situations in the Schiphol area.

4.2 Locked-in logistics further defined

The examples of paragraph 3 show that looking at the collective level (the Schiphol area) gives a completely different view than looking at the individual level (the individual companies or logistics chains). The Schiphol area can not simply be seen as the sum of the individual parts or elements. In essence, the total is different than the sum of its individual elements [Kramer; 1987]. In paragraph 3, the logistics sub-optimizations were seen at the collective level. If you want to change such a sub-optimization and you experience that it seems to be fixed, than you have to solve a collective locked-in logistics situation.

Locked-in logistics can also occur within individual companies or individual logistics chains (the individual level). In that case, at an individual company, there is a logistics sub-optimization that seems to be fixed. We will call this an individual locked-in logistics situation. KPMG-BEA and Kunkels did research about individual locked-in logistics and logistics chains [AVV; 199] [Kunkels; 1998]. Below follow two examples of possible individual locked-in logistics situations in the Schiphol area.
Example 1
In the Schiphol area, the airport related companies are often clustered and located along transport corridors [Kramer; 1998, p 11-14]. The exit of or the entrance to the motorway nearby a growing cluster of companies will have to absorb more and more traffic. This also means more merging traffic on the motorway. This growth of merging traffic can lead to congestion problems. A company location, chosen because of good accessibility, can turn, via this process, into a location with bad accessibility. A transport company that finds itself in this situation is dealing with a logistics sub-optimization. Nevertheless the company will not move to another location because of high switching costs.

Example 2
Company X is located close to Schiphol because of the short travel- and transport times to many important cities in the world via Schiphol Airport. The capacity of Schiphol Airport is limited. At this moment the growth of aircraft movements at Schiphol Airport is 4% per year. This growth can, at the airport, lead to: delays, "no flight available", and strong competition between capacity for passengers or capacity for air cargo (this last point will only be true for air cargo transported with full-freighters). For company X this can turn out into long and unreliable travel- and transport times via Schiphol Airport. Company X is dealing with a logistics sub-optimization but will not move to another location because of high switching costs.

Summarized:
Locked-in logistics will be defined as an inefficient logistics situation that is experienced as fixed or impossible to change. Locked-in logistics can occur on a collective level and an individual level.
5. Preventing and overcoming locked-in logistics

5.1 Introduction

With respect to the logistics sub-optimizations and locked-in logistics situations in the Schiphol area, we will follow two lines of reasoning:

a) Preventing future logistics sub-optimizations and locked-in logistics situations by changing the spatial behavior of companies. This is directed at companies who are not yet located in the Schiphol area but are making plans for doing so. We will call this management of company locations. (See paragraph 5.2)

b) To reduce or overcome current locked-in logistics situations by changing the logistics behavior of companies. We will call this management of modes of transportation. (See paragraph 5.3)

We can study the spatial and logistics behavior of companies from the perspective of the physical infrastructure. Via that perspective we will see: buildings, roads, airports, cars, planes, and traffic jams. We also can study the spatial and logistics behavior of companies from the perspective of the social infrastructure. Via that perspective we will see: procedures, institutions, networks, organization structures, rituals, paradigms, structures of knowledge to interpret information and to produce new knowledge. The decision making process is part of the social infrastructure. The physical and social infrastructure are connected, the one affects the other and the other way around. We will aim at influencing the physical infrastructure via the social infrastructure (see paragraph 5.4).

5.2 Spatial behavior of companies

Why does company X choose Y for it's place of business? The spatial behavior of companies is the result of [Lambooy; 1997, p 9-29]:

- **Company-internal factors:**
  This are factors as: strategy, organization, production, finance, logistics, and marketing.

- **Company-external factors:**
  This are factors as: clients, competitors, availability of employees, the services offered by Schiphol Airport, and the government. With government we mean national government (taxation, labor law, economic stability, infrastructure, and education) as well as local government (industrial sites).
A revaluation of company-internal factors will lead to a revaluation of company-external factors. So, company-internal factors and company-external factors are linked.

Decision making processes at the level of local government have a huge impact on the national spatial development. Frieling makes the point that the national spatial development in The Netherlands is determined at the level of local government [Frieling; 1999]. The allocation of a specific company to a specific location takes place at the level of local government. For studying the spatial behavior of companies, we will focus at: decision making processes within the company, decision making processes within the logistics chain, and decision making processes at public or private actors with respect to the allocation of space for company locations.

5.3 Logistics behavior of companies

What to do in case a locked-in logistics situation exists? We can reduce or overcome a current locked-in logistics situation (locked-in logistics at the collective level as well as locked-in logistics at the individual level, see paragraph 4.2) by changing the logistics behavior of companies. But, the problematic characteristic of a locked-in logistics situation is that it is experienced as fixed or difficult to change. The logistics behavior of companies is the result of:

- company-internal factors and,
- company-external factors.

Within the entire field of logistics behavior, our attention is mainly directed at: which modes of transportation are chosen, why, and can we change that. The company location (company-external factors) has a huge impact on the logistics behavior of the company (company-internal factors). A hypothetical example:

Suppose, Schiphol Airport is going to handle 25% more passengers. For DC X (Distribution Center), located close to Schiphol, this means a change in its environment or company location (company-external factors). On a passenger flight, often cargo is transported in the belly of the airplane as well. So, increasing the capacity for passengers automatically means increasing the capacity for air cargo. Airlines will try to get an efficient load factor for air cargo. Therefore DC X might be asked whether it is interested in sending a shipment via Schiphol for a very cheap price. Normally DC X sends that specific shipment via seaport Rotterdam, but now DC X decides for Schiphol. So, DC X discovers to use time and speed to create a new customer service element and to be more competitive. (In this context, speed is positive).
DC X discovers also that air cargo can be used as safety-net for all kinds of logistics problems. After some time, DC X converts to a different logistics behavior. Fast physical distribution via Schiphol becomes the general solution for logistics failures. (In this context speed is negative, it stands for bad logistics management). When this new logistics behavior is embedded, DC X feels that a modal shift (less air cargo via Schiphol and more shipments via seaport Rotterdam) is impossible. DC X isn't able to see anymore that the total lead-time offers much more possibilities for time-reduction than only fast physical distribution via Schiphol. For example, many other important possibilities for reducing lead-times are offered via better coordination of the activities along the goods flow.

We saw that the company location may have a huge impact on the logistics behavior of a company. Also the other way around. The logistics behavior may have a huge impact on how the company location is valued. An other hypothetical example:

DC X from the last example decides to implement a new information system and to work on a better coordination of the activities along the goods flow. This functions well. DC X discovers that air cargo as well as the expensive company location close to Schiphol are not needed anymore to meet the required customer service levels. Therefore, DC X would like to move to another company location. Now an individual locked-in situation can occur. Changing the logistics behavior of a company can reduce or overcome a collective locked-in logistics situation and result in an individual locked-in logistics situation.

Some conclusions:
- company-internal factors and company-external factors are linked,
- collective and individual locked-in logistics situations are linked,
- management of company locations and management of modes of transportation are linked.

The same decision making processes that are important for studying the spatial behavior of companies are important for studying the logistics behavior of companies. So, for studying the logistics behavior of companies, we will focus at: decision making processes within the company, decision making processes within the logistics chain, and decision making processes at public or private actors with respect to the allocation of space for company locations.
5.4 The decision making process and locked-in logistics

Now we will explore three theoretical approaches from the field of decision making to study locked-in logistics. Each of these approaches gives a different and interesting vision on locked-in logistics and decision making. Below, the three theoretical approaches are described in short.

The Prisoners Dilemma:
The conflict between individual and collective interest, is at the basis of the Prisoners Dilemma. Players in a Prisoners Dilemma game find themselves caught in a "catch 22" situation in which they are done by their rational calculations. Even though they are both better off if they cooperate, the logic of a dominant strategy dictates that each player, in pursuing his own self ends, defects from cooperation. The Prisoners Dilemma game shows the paradoxical nature of individually rational but collectively irrational decisions. [Zagare; 1988]

The Prisoners Dilemma offers a perspective to study collective locked-in logistics situations. The theory of the Prisoners Dilemma fits in with the concepts "logic of consequences" (the attention is aimed at optimizing the individual situation) and "logic of appropriateness" (the attention is aimed at optimizing the collective situation) [March; 1994]. By Teisman these concepts are translated in respectively "logic of consequent behavior" and "logic of interactive action" [Teisman; 1997, p 30].

Path dependence:
David defines path dependence as follows: "A path-dependent sequence of economic changes is one of which important influences upon the eventual outcome can be exerted by temporally remote events, including happenings dominated by chance elements rather than systematic forces" [David; 1985, p 332]. It is said that economic development is locked in a historical path and that departure thereof is not possible. Small and at first sight unimportant decisions can create determined conditions for future developments. Think about concepts as timing, sunk-costs, and switching costs. Path dependence offers a perspective to study collective and individual locked-in logistics situations.
The interaction perspective:
From this perspective, it is assumed that reality is constructed in social processes. People exchange their individual reality through interaction. In attempting to achieve common knowledge, a common reality is socially agreed upon. No reality exists in a social vacuum. Reality is not only constructed, but also continually reconstructed and changed in processes of ongoing interaction. There is no objective reality, only social realities which unfold in interaction.

Definitions of reality are constructed, reconstructed and changed in interaction. The definitions of reality are the cognitive dimension ("what"), the interaction patterns between social actors are the social dimension ("who" and "how"). "How" means the interaction rules. The "what", "who", and "how" are connected to each other. The central point for analyzing the process of change is the connection between social and cognitive aspects of interaction. Social change always arises from confrontation with variation. Interaction that always takes place within the same group of actors, when using the same definitions of reality, and when using the same interaction rules, will lead to circular processes. In this kind of situations only small changes are possible. Confrontation with variation can be the result of introducing a new definition of reality, a new actor, or a new interaction rule. Reducing variation can lead to fixation with respect to "what", "who", and "how". This fixation will lead to fixed patterns of behavior, fixed rules, or fixed conventions. [Termeer; 1993, p 11-41]

This fits in with March. March makes the point that "rule following" is an important part of decision making processes [March; 1994, p 57-102]. The interaction perspective can be applied for studying collective and individual locked-in logistics situations.
6. Conclusions and further research

If we focus on the Schiphol area, and within that area on, freight transport, modes of transportation and company locations it seems that some logistics sub-optimizations are "locked-in" in our transport system (locked-in logistics). How to prevent that and how to overcome that?

This paper offers a first step towards a framework for analyzing, preventing and overcoming locked-in logistics. Via the social infrastructure we try to understand and to influence the physical infrastructure. The decision making processes are part of the social infrastructure. The decision making process offers interesting viewpoints for analyzing and explaining locked-in logistics.

Our final goal is to offer a framework for solving locked-in logistics situations within the Schiphol area via improving: the decision making processes within companies, the decision making processes within the logistics chains, and the decision making processes at public or private actors with respect to the allocation of space for company locations.
References


• BCI/NEI, "Ruimtelijke-economische verkennning van de Toekomstige Nederlandse Luchtaart Infrastructuur" (Spatial-economical outlook, the future of the Dutch aviation-infrastructure), 1997. (In Dutch).

• Bouwens, A.M.C.M., M.L.J. Dierikx, "Tachtig jaar Schiphol; op de drempel van de lucht" (Eighty years Schiphol Airport; at the doorstep of the air), Sdu Uitgevers, Den Haag, 1996. (In Dutch).


• Frieling, D.H., "Het metropolitane debat" (The metropolitan debate), in: Kenmerken, 6, 1 februari 1999, blz. 4-7. (In Dutch).


• Kunkels, B.H.R., "Vooruitgang is iets anders dan voortrazen op de ingeslagen weg" (Progress is something else than proceeding fast along the same line), doctoraalscriptie, Universiteit van Amsterdam, Faculteit der Economische Wetenschappen en Econometrie, 1998. (In Dutch).


• Oosterhaven, J., F.J. Sijtsma, "Over de directe en indirecte economische betekenis van de luchthaven Schiphol" (About the direct and indirect economical meaning of Schiphol Airport), hoofdstuk 10 in de bundel "Transport en welvaart; vervoer in vogelvlucht", Stichting REG, Groningen, 1998. (In Dutch).

• Teisman, G.R., "Sturen via creatieve concurrentie; een innovatie-planologisch perspectief op ruimtelijke investeringsprojecten" (Steering via creative competition; a perspective for spatial projects), 1997. (In Dutch).

• Termeer, C.J.A.M., "Dynamiek en inertie rondom mestbeleid" (Dynamics and inertia at the slurry problem policy), 1993. (In Dutch).

Modeling Poiseurges in Tunnel Stations

TRAIL Research School, Daily December 1990

Author:

R.W. Opper

Title of Thesis:

"Modeling Poiseurges in Tunnel Stations"

Presented at: TRAIL Research School, Daily December 1990

Committee:

R.W. Opper, Chair

H. van den Berg

T. Hoekema

R. van den Berg

Chair of Committee:

R.W. Opper

Date of Defense:

December 1990

Copy of Defense:

R.W. Opper

10
Abstract

This paper deals with the outline of a pedestrian simulation model, especially for
transfer stations. The model is an aid to visualize and to quantify effects of
platform design choices on the pattern of passenger flows through a
transfer station. This helps to reveal bottlenecks that may cause congestion and
long waiting times, both contributing to long transfer times.

The model also enables an analysis of effects of alterations of the various, public
transport systems. Some critical aspects were identified at some of the stations of
the Dutch railway network. The input for the tool is a
set of parameters and datasets of the pedestrians and the
infrastructure, and the dynamics in the model is
represented due to changes in system (arrival and departure of trains, buses and
passengers).

1 Introduction

2 Modular structure of the tool

3 Modelling pedestrians
  3.1 Requirements
  3.2 Levels of modelling

4 Application of the pedestrian model
  4.1 Aspects of implementation
  4.2 Input for the model
  4.3 Output of the model

5 Status of the project

6 Practical application

7 Conclusions

References
Abstract

This paper deals with the outline of a pedestrian simulation model, especially for transfer stations. The model is an aid to visualise and to quantify effects of station layout design choices on the pattern of passenger flows through a transfer station. This helps to reveal bottlenecks that may cause congestion and long walking distances both contributing to long transfer times.

The model also enables analysis of effects of timetables of the various public transport systems (departure and arrival times of the services at transfer stations) on passenger flows in and through the transfer station (congestion, transfer times).

In addition, the effects of evacuations are visualised. The input for the tool are a station layout, the origins and destinations of the pedestrians and the timetables for the different (public) transport modes. With this information the tool can calculate the following data: walking directions, walking times, transfer characteristics, consequences for the infrastructure, and the dynamics in these quantities due to changing conditions (arriving and departing trains, buses and passengers).

First, the paper explains the application purposes of the model and its modular structure. Then, the requirements from the tool are specified, as well as the pedestrian model and the requirements from the input and the output. The paper concludes with a practical application of the tool and the progress of the research.
Chapter Title

The chapter title would be "The role of a beneficial communication model especially for..."

The chapter continues with text that is not legible due to the image quality. It discusses the importance of beneficial communication, possibly in a specific context such as education or mental health, and emphasizes the role of effective communication in fostering understanding and cooperation.

Emphasis is placed on the need for clear and supportive communication strategies, highlighting the importance of active listening and empathy in creating a positive and inclusive environment. The text also touches on the challenges faced in implementing such strategies and suggests potential solutions for overcoming these obstacles.
1 Introduction

Background

The traditional way of designing transfer stations is based on rules of thumb. These rules are based on many years' experience concerning the behaviour of passengers in transfer stations, but they only consider static situations. Until now, a scientific foundation for these rules has not yet been provided. However, even the dynamic behaviour of pedestrians needs to be taken into account to predict the consequences of a station layout design in practice.

Currently, timetable generation involves the application of approximations for the time people actually need to transfer. However, different types of passengers (elderly people, parents with children) need different transfer times. Adopting accurately estimated transfer times will reduce the probability for passengers to miss their connection, thus increasing the traveller's comfort and the timetable's reliability.

Purpose

In association with Holland Railconsult, a simulation tool is under development to estimate the walking times incurred by transferring passengers (mean and variance) and to visualise walking patterns inside transfer stations. Simulation studies with this tool may reveal the level of comfort for passengers in transfer stations.

Requirements

General traffic characteristics of pedestrians are adjusted to those of passengers in transfer stations and calibrated using Dutch empirical data in order to obtain an appropriate model for the Dutch situation. The model also describes how the transfer services are performed. Finally, an object oriented simulation model is developed and subsequently validated and calibrated by performing several case studies.

Structure of the paper

This paper gives a description of the simulation tool. First, the modular structure of the tool is illustrated. Then the requirements for the tool are specified. Section 5 pays special attention to the application of the pedestrian model and the input and output requirements. The paper concludes with the status of the projects and our intentions for the future.
2 Application purposes of the model

The simulation model can be applied to quantify the level of comfort of pedestrians while they are moving through the station and while they are waiting at the platform or in the hall. These situations can be simulated for existing stations, for extensions for existing stations and for stations under development.

Also, the transfer behaviour of passengers can be evaluated. This deals with the amount of passengers who succeeded in their transfer and with the waiting time of the transferring passengers. Causes of missed transfers can be the lay-out of the station-infrastructure, the timetable or the track-allocation. These data are part of the model to be able to analyse the effect on pedestrian flows. Different timetables and track-allocations can be compared, after which the best alternative can be chosen.

New in this simulation tool is the modelling of the dynamical behaviour of pedestrians. Existing models, like PEDROUTE, have a more static based assignment of pedestrians to the network. Moreover, the output of the model is fixed and different transport modes cannot be all part of the model. In our model, these modes are included, as well as the interactions between the vehicles and the pedestrians (especially to get on and off these vehicles). Even activities in the station, like buying tickets and pedestrian' behaviour in front of ticket offices, are taken into account.

In short, the tool in development is used to evaluate:
- the layout of existing transfer stations,
- effects of extensions and/or adaptations of existing transfer stations,
- layout alternatives for newly developed stations,
- changes in platform allocation and
- changes in the timetable.
3 Modular structure of the tool

The required tool describes both the behaviour of passengers and different (public) transport modes and the interaction between them. Figure 1 shows an overview of the modular structure of the tool.

![Modular Structure Diagram]

Figure 1: The modular structure of the tool (example)

The pedestrian module is currently under development and is called SimPedestrian. The module includes three theoretical behaviour models: the behaviour of pedestrians on infrastructure available for walking (walkways, stairs, elevators etc.), their behaviour on infrastructure available to perform activities (eg. buying tickets or a newspaper) and their behaviour in relation to other transport modes (eg. getting on and off a train).

Holland Railconsult has developed a tool named SimRail, which is currently used to perform railway simulation studies. SimRail will be used as the rail module in Figure 1. In SimRail both the rail infrastructure and the behaviour of vehicles on this infrastructure are modelled. Modules for other transport modes will also be developed, like SimBus and SimCar and especially for the Dutch case SimBicycle.

The tool offers an objective view on the level of pedestrians' walking comfort in a transfer station, depending on the quality of the infrastructure for pedestrians. The deviations from the scheduled arrival and departure times of vehicles also play an important role in the passenger process: arriving vehicles determine the moments of getting on and off the train and, indirectly, the waiting time of transferring passengers. Exact origins and destinations of the pedestrians are influenced by the allocated positions of the stopping vehicles. The station manager is responsible for this allocation process and can, beforehand, evaluate the effects of a platform change.
4 Modelling pedestrians

Modelling and visualising the dynamic behaviour of pedestrians, especially at railway stations is rather new. At transfer stations, people behave differently compared to normal circumstances (e.g. they may be in a hurry to catch their train or they are down because they have missed it). Moreover, the walking conditions in stations differ substantially from those outside (higher densities of people, wider range of walking purposes and speeds). Therefore, the specific behaviour of passengers in transfer stations needs to be studied.

4.1 Requirements

The tool should be able to handle large transfer stations of at least 100,000 m$^2$ area intended for walking purposes. Furthermore, it should be prepared for the presence of about 100,000 persons during the same period of time. The developing process starts with the construction of a tool containing only pedestrians, rolling stock and the interactions between pedestrians and rolling stock (i.e. getting on and off the train). Afterwards, the concept is used to produce a general architecture which can be extended with all other modes of transport. The modelling of walking behaviour of pedestrians may be based on general traffic flow theory (see Daganzo (1997) and May (1990)). In this project, the validation of the walking behaviour of pedestrians will be based on observations in Dutch railway stations (for example Utrecht CS).

4.2 Levels of modelling

Pedestrian behaviour is modelled in two ways: microscopic and macroscopic. Microscopic modelling means that each pedestrian is a single object and has interactions with other pedestrians (see e.g. Blue (1998) and Nagel (1998)). In a macroscopic model, streams of pedestrians move through the station network instead of single ones (see Hoogendoorn (2000)). With a macroscopic model, the level of comfort of parts of the network can be assessed immediately and congestion can be detected. Subsequently, microscopic simulations are used at these specific parts of the transfer station in order to determine those causes for the congestion, which cannot be found by macroscopic simulation.
5 Application of the pedestrian model

The model is used to quantify the effects of station layouts on pedestrians in a transfer station. Visualising these effects helps to identify and to understand the causes for the presence of bottlenecks, long walking distances and large transfer times. The visualisation is achieved in two ways: macroscopic (indicating the level of comfort of a walkway) and microscopic (related to the individual pedestrian). Furthermore, the tool helps to determine the effect disruptions during operation (delays and early arrivals of trains) may have on the transfer times of passengers.

In evaluating the resulting pedestrian flows, the tool is considered as a black box, containing a model which allocates pedestrians to a network (see Figure 2). Inputs for the allocation model are a network with different links and a table of pedestrian origins and destinations. The behaviour of the different types of pedestrians (tourists, workers, etc) is also taken into account.

During simulation, the tool performs the following activities repetitively:
- determining origins and destinations of the present pedestrians,
- assigning routes through the station network,
- calculating walking times based on the behaviour models and
- executing situation updates and dynamic evaluations.

The tool produces different walking routes used by the pedestrians. Also, walking times are calculated by the tool, in terms of means and standard deviations. From the number of pedestrians on a link, the mean available space for each pedestrian can be derived which, in turn, is an indication for the performance on the link. This quantity is expressed in a level of comfort. Low levels of comfort indicate the presence of congestion in the network (see Fruin (1971)).

![Figure 2: The pedestrian movement evaluation tool including inputs and outputs](image-url)
5.1 Aspects of implementation

The theoretical behaviour models need to be modified for transfer station conditions. To achieve this, the following aspects are subsequently implemented in the tool:

- Infrastructure for pedestrians to walk on. Infrastructure elements can be classified as walkways (including stairs and passages), escalators or platforms.
- Infrastructure to service pedestrians. Initially, only ticket machines, ticket offices and kiosks are taken into account. These are modelled as areas, in front of which pedestrians queue up while waiting for their turn. These queues form, at their turn, obstacles for other pedestrians. The service time of each customer is drawn from a specified probability distribution.
- Route choices allowing pedestrians to walk from their origins to their destinations. The network consists of linked walkways, escalators and platforms. An algorithm is used to search paths through the network. Each pedestrian has its own shortest path in time. The duration of a route depends on the number of pedestrians present.
- Interaction of the pedestrians with the rolling stock. That is, getting on and off the trains. For each pedestrian the boarding time is drawn from a distribution, with its parameters depending on his or her age and physical condition. The parameters are calibrated by observations in practice. The same procedure is followed for the time used by a pedestrian to get off the train.

5.2 Input for the model

The user has to specify the following aspects in order to make a simulation model.

Network data:
- Railway infrastructure. The infrastructure consists of block sections, switches and platform tracks used by trains to arrive at and depart from the stations.
- Infrastructure for walking purposes in the transfer station. The infrastructure consists of walkways, passages, platforms, stairs, escalators and elevators.
- Infrastructure for services available to pedestrians. These services may be station related (e.g., ticket-offices, stamp and ticket machines) or not (e.g., cafeterias, self-service snack-bars, kiosks and small bookshops).
Behaviour data:
- Rolling stock. Characteristics of rolling stock influencing the time pedestrians need to get on and off the trains. Also the dynamic train characteristics are to be specified.
- Pedestrian characteristics. Each type of passenger is characterised by range of ages and travel motives. The corresponding walking speed is derived from these characteristics.

Origin and destination data:
- Timetable. The tool generates the train arrivals and departures either according to a timetable or randomly according to an exploitation pattern.
- Origins and destinations of pedestrians. For each train series the number of passengers getting on and off the train is to be specified. Furthermore, the user should define how the passengers are distributed both in the train and on the platform.

5.3 Output of the model

The tool produces the following output:
- trip information for each train (punctuality and achieved running speed etc.),
- trip information for each pedestrian (routes and walking times etc.),
- transfer information (transfer time and waiting time of a transferring pedestrian etc.),
- station infrastructure information (walkway conditions, service times and congestion areas etc.).
6 Practical application

This example shows how the simulation tool can be used for practical situations. In this case we point out that no problems occur in the level of comfort of pedestrians when a train arrives according to the timetable. When a train arrives too late, the level of comfort decreases significantly. The modelled situation appears at Utrecht Central Station at platform 4B when the train from Arnhem arrives.

This simulation is executed with the simulation tool, developed during my graduation (see Daamen (1998a)). The model contains a microscopic assignment model for pedestrians. To assign speeds to pedestrians, empirical data from literature have been used (see Fruin (1971)). However, the overtaking behaviour of pedestrians comes up not so good, but that can be solved by a division of the infrastructure in many small parts, which can be connected. Moreover, the model is quite slow, which makes it hard to perform simulations for large transfer stations.

The infrastructure consists of a platform with a length of 200 m. An entrance/exit is located at the left of the platform through which pedestrians enter and leave. The train arrives from the right and dwells at the beginning of the platform (at the right in the drawing). The train and leaves in the same direction as where it came from. A model of these infrastructure is made in Figure 3.

It is clear that passengers getting off the train and walking to the exit will be hindered by passengers getting on the train and vice versa. According to the timetable the train arrives 15 minutes before its planned departure at the station. The passengers getting off the train have already left the platform, before the boarding passengers arrive. But sometimes trains do arrive too late, so passengers getting off the train will hinder boarding passengers. This means that more passengers are present at the same platform, through which the level of comfort decreases. The densities at the left part of the platform are recorded and are shown in Figure 4 and Figure 5.
It appears from these two figures that when the train arrives too late (15 minutes after the planned time of arrival) the level of comfort is below level B during 2 minutes, while in the situation when the train is in time, this is only 30 seconds the case. In Figure 4 two peaks can be distinguished: one peak for the passengers getting off the train and the other peak for boarding passengers. In Figure 5 these peaks coincide, which causes the overload for the infrastructure. This case shows that when a train arrives according to the timetable, no overload appears on the infrastructure, while at a delay the level of comfort decreases significantly by simultaneous getting on and off the train.

Besides, this case demonstrates that a clear impression can be given of the densities and the level of comfort in a transfer station, given a timetable.

---

TRAIL Research School, December 1999
7 Progress of the research

In this paragraph an overview of the development of the simulation tool is given. Then, findings from performed research are presented and followed by the scientific research, which has to be carried out.

7.1 Development

As mentioned before, the tool consists of several modules. The first version of the tool will only contain the modules for pedestrian behaviour and rolling stock behaviour. The latter module has already been developed by Holland Railconsult and is called SimRail. The pedestrian module is under development. It contains a theoretical behaviour model for pedestrians. Different theoretical models are already available. They are compared to determine which model reflects the walking behaviour at transfer stations most accurately (see Daamen (1999)).

A software architecture has been designed to combine the different modules of the tool. For an initial version of the tool, the architecture will be implemented and the pedestrian module has to be completed. According to the current schedule, this will be the case in December 1999.

During the following year, the tool will be validated using observations in Dutch railway stations. Data collection for this purpose is planned for the beginning of the year 2000. Afterwards, small projects may be performed to help validating the tool in co-operation with clients on the railway market. Simultaneously, other modules are designed and implemented in order to provide additional functions (eg SimBus and SimCar).

7.2 Findings

Few literature is available about the behaviour of pedestrians, and specifically the behaviour of pedestrians in transfer stations. Moreover, this literature dates from the sixties and the seventies. Only recently, articles have been published on dynamic traffic flow models for the simulation of pedestrian behaviour.

To model pedestrian, a model has already been developed (PEDROUTE), which does not provide in flexible output and dynamic allocation of pedestrian to a network. Also, the input of a timetable for Public Transport is hard. The alternative, the traffic allocation model AIMSUN is not able to model pedestrians. Using a general simulation tool (like ARENA or MODSIM) all required functionalities can be included and statistics can be tuned to the requirements of the user (see Daamen (1998b)).
7.3 Scientific research

Different traffic flow models have been compared to microscopically modelled pedestrians. We have made simulation models of these theoretical models, with which tests have been performed. These models have resulted in fundamental diagrams for traffic behaviour. These fundamental diagrams have been compared, whereupon the Cellular Automata model appeared to be the best (see Daamen (1999a)).

Research is carried out on the behaviour of pedestrians on escalators. Fundamental diagrams are the basis for this behaviour, as they are valid for car traffic and pedestrian traffic in passages. In this research escalators are compared to stairs and the effects of walking on escalators are examined. Also, the influences of the speed of an escalator and its width are considered.

As has been stated, little research has been done on behaviour of passengers in Dutch transfer stations. For verification and validation of the tool, quantitative data are necessary on fundamental parameters like speed, density and intensity, for different kinds of infrastructure (passages, stairs, etc). For this purpose, observations are carried out in a Dutch railway station, preferably Utrecht CS. A fact-finding research has already been started, which has indicated suitable locations to make these observations and the planning of the observations (what has to be measured, what infrastructure, which period of the day) (see Daamen (1999b)).
8 Conclusions

In this paper we have presented an outline of the simulation tool for visualising pedestrian flows in transfer stations. Measuring transfer times of pedestrians makes it possible to objectively evaluate level of comfort for pedestrians and, with these results, compare alternative station layouts. Also, effects of different timetables and different platform tracks allocations on pedestrian flows are visualised. The tool simulates various modes of (public) transport, pedestrians and the interaction between them.

Pedestrians are modelled in two ways: macroscopic and microscopic. In the microscopic model, all pedestrians are different objects, interacting with each other. The macroscopic model is concerned with aggregate pedestrian flows.

To evaluate the infrastructure of the entire station, the macroscopic model is used to calculate levels of comfort. If the level of comfort is insufficient, the corresponding part of the infrastructure is modelled microscopically in order to reveal the causes for the congestion.

At this stage of the project, the pedestrian behaviour model and the railway traffic model are implemented. In the near future, the tool will be extended to allow simulations with other transport modes and inside other pedestrian areas, like Central Business Districts, stadiums and airports.

Acknowledgements

This publication is a result of the research programme Seamless Multimodal Mobility (SMM), carried out within the TRAIL Research School for Transport, Infrastructure and Logistics, and financed by Delft University of Technology in cooperation with Holland Railconsult.
References


Conclusions

This publication is a result of the research programme "Transport Behaviour Models and Applications (TRAM)". It was part of the "Teacher-Training Programme on Transport Infrastructure and Logistics" and financed by Delft University of Technology in cooperation with regional authorities of the Netherlands.

Acknowledgements

References
On the Optimisation of Buffer Times in Transport Networks

TRAIL Research School, Delft, December 1999

Author
Ir. R.-J. van Egmond
Information Technology and Systems, Delft University of Technology

Thesis supervisors
Prof. dr. G.J. Olsder
Prof. dr. M.F. Dekking
Information Technology and Systems, Delft University of Technology
On the Optimization of Buffer Times in Transport Networks

TRAI Research School Draft December 1996

Author

[Signature]

Trustee: Professor G.T. Oates

PAX Institute of Technology

Engineering Technology and Systems Design Department of Technology
## Contents

**Abstract** ii

1 **Introduction** 1

2 **Propagation of initial delays** 2
   2.1 Event graphs and maxplus algebra 2
   2.2 Simulation 3
   2.3 Analysis 4
   2.4 Buffer times 4

3 **Optimal buffer times** 5
   3.1 Acyclic networks 5
   3.2 Cyclic networks 6

4 **Conclusions and further research** 8

References 9

Appendix A: A circuit in the active Ip-constraints 10

Appendix B: The Floyd-Warshall algorithm 11
Abstract

The event graph approach is used to design optimal timetables for transport networks. The objective is to tackle the trade off between throughput and punctuality; find buffer times such that the level of punctuality is satisfactory and the throughput is maximised. This paper provides a method to evaluate a transport system when the buffer times and distributions of initial delays are given. Also the problem of finding optimal buffer times is partially solved; one part of the optimisation problem can be solved by linear programming. For the other part a "trial and error" method is suggested.
The main duty of the farmer is to balance the needs of producing food and energy. The farmer's role is to ensure the land is used efficiently and sustainably. This means that the level of biodiversity and environmental protection is maintained and the farmer's actions contribute to the health of the ecosystem. The farmer must also take into account the needs of the local community, ensuring that the food produced is both healthy and affordable. The farmer's role is to ensure that the food produced is not only nutritious but also culturally significant. The farmer must also consider the impact of farming on the environment, ensuring that the actions taken do not harm the natural world. This is achieved through the use of sustainable farming practices, such as crop rotation and the use of organic fertilizers.
1 Introduction

The scope of this paper is to design timetables for transport networks that make optimal use of the available resources. The objective is to maximise the throughput of the system (i.e. the number of vehicle movements scheduled per time unit). The throughput is bounded by the required quality of operations: if the throughput is too high, the punctuality (which is especially important in transport of persons) may drop below an acceptable level. Hence, the throughput of a timetable should be maximised under the condition that the level of punctuality remains satisfactory.

In this paper, we investigate timetables with the aid of event graph theory. The power of event graph theory lies in the performance evaluation of cyclic schedules. In cyclic schedules, the throughput is given by the cycle time of the pattern. In order to make optimal use of the available resources, the cycle time should be minimised. The cycle time and punctuality are related by buffer times: increasing buffer times will improve the punctuality and worsen the cycle time. In order to find optimal buffer times, we first investigate how delays propagate when buffer times are given. Then we relate delay propagation to punctuality when distributions of initial delays are given. Finally, we are interested in finding buffer times that fulfil constraints on the punctuality and minimise the cycle time.

This paper is organised as follows: in Section 2 we give a brief overview of event graph theory and its application to delay propagation. This work has been presented earlier in [3]. In Section 3 one can find the main results of this paper: the connection between delay propagation and punctuality and the problem of finding optimal buffer times. The conclusions of this paper are summarised in Section 4.
2 Propagation of initial delays

2.1 Event graphs and maxplus algebra

First we give a brief introduction in the theory of event graphs. An extensive discussion of event graphs can be found in [2].

In the theory of event graphs we study systems of the following form:

\[ x_j(k+1) = \max_{i=1,...,n} \{ x_j(k) + a_{ij} \}; \quad i = 1,...,n; \quad k = 0,1,..., \]

or, in matrix/vector notation:

\[ x(k+1) = A \otimes x(k); \quad k = 0,1,..., \]

where \( x(0) \) is the initial condition vector. This equation allows the following physical interpretation: Consider a transportation network for which a schedule has to be designed such that for each node vehicles leaving from this node have the same departure time. Let \( x_j \) denote the departure time at node \( j \) and \( a_{ij} \) the running time from node \( j \) to node \( i \). In order to provide a transfer (for passengers or cargo), vehicles have to wait for one another, i.e. the \( k+1 \)th departure at node \( i \) takes place when all vehicles of the \( k \)th departure travelling from node \( j \) to node \( i \) have arrived at node \( i \). Hence, the evolution of the departure times is given by equation (1). In this example vehicles wait for one another in order to provide a transfer. However, event graphs can also be used to model vehicles that wait for one another because of the common use of infrastructure.

If no connection between nodes \( j \) and \( i \) exists, we set \( a_{ij} = -\infty \) and add the convention that \( \max\{a, -\infty\} = \max\{-\infty, a\} = a, \forall a \in \mathbb{R} \).

Equation (1) can be viewed as a linear system by writing the \( \max \) operator as \( \oplus \) and the \( + \) operator as \( \otimes \) (the \( \oplus \) and \( \otimes \) around these new operators indicate that we change from the conventional algebra to another algebra: the maxplus algebra):

\[ x_j(k+1) = \bigoplus_{1 \leq j \leq n} x_j(k) \otimes a_{ij}; \quad k = 0,1,..., \]

or, in matrix/vector notation:

\[ x(k+1) = A \otimes x(k); \quad k = 0,1,..., \]

where \( A \) denotes the matrix with entries \( a_{ij} \) and where matrix multiplication is defined in the similar way as in conventional algebra. We say: synchronisation constraints are linear in the maxplus algebra.

Next, we question whether an initial vector \( x(0) \) exists such that the system behaves in a regular way, that is:

\[ x(k+1) = \lambda \otimes x(k); \quad k = 0,1,..., \]
for some $\lambda \in \mathbb{C}$. The interpretation in terms of the transport network is that for each node the interval between successive departure times is constant. Another way to state this problem is to find a vector $v$ and a scalar $\lambda$ such that

$$A \otimes v = \lambda \otimes v.$$  

We call $v$ an eigenvector and $\lambda$ an eigenvalue of matrix $A$, which can be interpreted as a regular timetable and the cycletime of the system.

The most important result on the eigenvalue problem (2) is that if matrix $A$ is irreducible then the eigenvalue $\lambda$ is unique and equals the circuit mean (i.e. the sum of the running times on a circuit divided by the number of nodes on that circuit) maximised over all circuits. The cycletime of a system is thus determined by circuits with maximal circuit mean, which are therefore called critical circuits. Furthermore, efficient algorithms exist to calculate eigenvalues, eigenvectors and critical circuits.

In the sequel we assume that matrix $A$ is irreducible, which is equivalent to the assumption that the underlying network is strongly connected, i.e. for each pair $(i,j)$ there exists a path from node $j$ to node $i$.

### 2.2 Simulation

Consider a timetable determined by an eigenvector $v$ of matrix $A$:

$$x(k) = \lambda^k \otimes v.$$ 

Suppose that at time $k = 0$ there are some delays $d(0) \geq 0$ (with $d(0)$ an initial delay at node $i$ and time $k = 0$), resulting in a vector of disturbed moments of departure $x'(0) = x(0) + d(0)$. Since we assume that all connections remain the same, the time evolution of $x'$ can be found by applying matrix $A$ and timetable $x$:

$$x'(k+1) = A \otimes x'(k) \oplus x(k+1).$$ 

The difference between $x'$ and $x$ determines the propagation of delay $d(0)$:

$$d(k) = x'(k) - x(k).$$

This way of calculating the propagation of delays is thus merely a matter of simulation. Also if matrix $A$ is replaced by random matrices $A(k)$, i.e. if the running times are considered as stochastic variables, this way of simulation can be used.
2.3 Analysis

In this section we investigate which nodes will be disturbed by a particular delay. Considering two arbitrary nodes we question how large a delay in one of these nodes must be to affect the other node. This leads us to the following definition:

\[ m_{ij} \] is the maximum delay at node \( j \) that does not reach node \( i \).

The values \( m_{ij} \) can be obtained using the following lemma:

**Lemma 1:** Let \( A \) be an irreducible matrix with eigenvector \( \mathbf{v} \) and eigenvalue \( \lambda \), then

\[ m_{ij} = v_i - v_j - \max_{P \in P_{ij}} \sum_{(k,j) \in P} (a_{kj} - \lambda), \tag{3} \]

where \( P_{ij} \) denotes the set of all paths from node \( j \) to node \( i \). We refer to [3] for the proof.

2.4 Buffer times

In order to increase the punctuality of the transport system, buffer times \( b_{ij} \) are added to decrease delay propagation. The system including buffer times is described by a new matrix \( A' \) with \( a'_{ij} = a_{ij} + b_{ij} \).

The new timetable and cycle time follows from the eigenvector and eigenvalue of \( A' \):

\[ A' \otimes \mathbf{v}' = \lambda' \otimes \mathbf{v}'. \]

Again, we are interested in the maximum delay in node \( j \) that does not reach node \( i \), denoted by \( m'_{ij} \).

**Lemma 2:** Let \( A' \) be an irreducible matrix including buffer times, then

\[ m'_{ij} = v'_i - v'_j - \max_{P \in P_{ij}} \sum_{(k,j) \in P} (a_{kj} - \lambda'). \tag{4} \]

Again, we refer to [3] for the proof. Notice that the lacking of the apostrophe of \( a_{ij} \) in Lemma 2 is not a misprint. On the contrary, it is essential that the buffer times disappear in the synchronisation constraints when calculating the propagation of delays.
3 Optimal buffer times

3.1 Acyclic networks

We consider an event graph, but now we assume the network to be acyclic; instead of a dynamic model describing the \( x^\text{th} \) departure from a node we consider a static model with the departures as separate variables. We use the following notations:

- \( x_i \) moment of departure at node \( i \),
- \( v_i \) moment of departure at node \( i \) according to the timetable,
- \( a_{ij} \) running time from node \( j \) to node \( i \),
- \( D_i \) initial delay at node \( i \).

Here, we write stochastic variables in capitals and deterministic variables in normal letters. The buffer times are encapsulated within the timetable: the buffer time between node \( j \) and node \( i \) equals \( v_i - (v_j + a_{ij}) \), i.e. the departure time at node \( i \) minus the arrival time from the vehicle coming from node \( j \) at node \( i \). The timetable \( v \) might be the result of spectral analysis of the dynamic model (cf. (2)). The behaviour of the system can be written as:

\[
X_i = \max_j \{ v_i + D_j, \max_j \{ x_j + a_{ij} \} \}.
\]  

In order to obtain an explicit expression for \( X \), we replace \( x_j \) in the righthand side of (5) by the whole righthand side of (5) and we keep doing this until \( X \) vanishes in the righthand side (this is guaranteed by the acyclicity of the network). We end with the following explicit expression for \( X \):

\[
X_i = v_i + \max_j \{ D_j - m_{ij} \},
\]

where

\[
m_{ij} = \begin{cases} v_i - v_j - \max_{p \in P_i} \sum_{k \neq p} a_{ik} & i \neq j \\ 0 & i = j. \end{cases}
\]

Again, we can interpret \( m_{ij} \) as the maximum delay at node \( j \) that does not reach node \( i \). Next, given the distributions of initial delays, we are interested in distributions of departure times. Assuming that the initial delays are independent we can obtain from (6):

\[
P(X_i - v_i \leq x) = \prod_j P(D_j \leq x + m_{ij})
\]
Assuming that vehicles never depart earlier than \( v_i \) is equivalent to assuming that \( P(D_i < 0) = 0 \). By this assumption we can interpret \( X_i - v_i \) as the waiting time of a passenger who planned to depart at \( v_i \) from node \( i \) and has to wait \( X_i - v_i \) time units before the vehicle actually departs. So, for a given timetable we can evaluate the punctuality via (8), but how can we find a timetable that meets given constraints on the punctuality? This problem can be split up in two parts: First, we search for appropriate constraints on the delay propagation \( (m_{ij}) \) that guarantees satisfactory punctuality. Second, we search for a timetable that meets given constraints on \( m_{ij} \). The first part is formulated as:

**Problem 1:** Given constraints on the distributions of the waiting times \( X_i - v_i \), translate these constraints in terms of the values \( m_{ij} \).

This is a difficult problem for which we do not know yet how to solve it in an elegant way. A rough way is just to try many sets of \( m_{ij} \)-values until the resulting waiting time distributions are satisfactory.

Once we have found appropriate constraints (lower bounds) on \( m_{ij} \), we search for a timetable \( v \) such that these constraints are met. For sake of efficiency it is natural to try to meet the contraints with a timetable which is 'as fast as possible', for instance a timetable that minimises the latest departure time.

**Problem 2:** Given an acyclic network and lower bounds on \( m_{ij} \), find a timetable \( v \) which obeys these constraints and minimises the latest departure.

This problem can be written as the following optimisation problem:

\[
\begin{align*}
\min_{v, T} & \ T \\
\text{s.t.} & \ v_i - v_j \geq m_{ij} + \max_{p \in \pi_i^{(k)}} \sum_{(k,j) \in p} a_{ij} \quad \forall i, j \\
& \ v_i - v_j + T \geq 0 \quad \forall i, j
\end{align*}
\]

where \( m_{ij} \) is a lower bound for \( m_{ij} \) and \( T \) is the makespan (difference between first and latest departure). We assume that \( m_{ij} \geq 0 \). By this assumption, contraints (10) incorporate the fact that vehicles wait for one another, since the constraints \( v_i \geq v_j + a_{ij} \) for each arc \((j,i)\) are automatically fulfilled then. Problem (9)-(11) is a linear programming problem which can be solved efficiently by standard solvers. Notice that the problem of minimising the latest arrival instead of the latest departure can be formulated along the same lines.

### 3.2 Cyclic networks

Reconsider a cyclic network and recall that for a given cyclic timetable the delay propagation is given by (3). Being interested in the "fastest possible" timetable for a cyclic network, a natural objective is to minimise the cycletime:

\[ \min T \]

\[ \text{s.t.} \ v_i - v_j \geq m_{ij} + \max_{p \in \pi_i^{(k)}} \sum_{(k,j) \in p} a_{ij} \quad \forall i, j \]

\[ \ v_i - v_j + T \geq 0 \quad \forall i, j \]
Problem 2*: Given a cyclic network and lower bounds on $m_{ij}$, find a timetable $v$ which obeys these constraints and minimises the cycletime.

This problem can be written as the following optimisation problem:

$$\min_{v, \lambda} \lambda$$

s.t. $v_i - v_j - \max_{p \in P_{ij}} \sum_{(k,j) \in p} (a_{kl} - \lambda) \geq m_{ij} \quad \forall i, j$. \hspace{1cm} (13)

Problem (12)-(13) can be written as a linear programming problem by writing the constraints in (13) for each path separately. However, the number of constraints (the number of paths) is combinatorial in the size of the network. Therefore, when solving (12)-(13) we consider the relaxed problem with constraints on the arcs only (this guarantees a bounded solution) and we add violated constraints of the original problem to the relaxed problem until an optimal solution of (12)-(13) is found.

As in event graph theory, we may characterise the cycletime in terms of critical circuits. Consider a solution $(v^*, \lambda^*)$ of problem (12)-(13) and a set $c^*$ consisting of active constraints over paths $P_{ij}$ that form a circuit (such a set always exists, see appendix A). Summation of these constraints leads to:

$$\sum_{P_{ij} \in c^*} (v_i - v_j) - \sum_{P_{ij} \in c^*} \sum_{(k,j) \in P_{ij}} (a_{kl} - \lambda^*) = \sum_{P_{ij} \in c^*} m_{ij}$$

$$\Rightarrow |c^*|\lambda^* = \sum_{P_{ij} \in c^*} m_{ij} + \sum_{(k,j) \in P_{ij}} a_{kl}$$

$$\Rightarrow \lambda^* = \frac{\sum_{P_{ij} \in c^*} m_{ij} + \sum_{(k,j) \in P_{ij}} a_{kl}}{|c^*|}$$ \hspace{1cm} (14)

where $|c^*|$ is the number of arcs in $c^*$. Furthermore, if another set $c$ contains an inactive constraint the equality in (14) becomes a strict inequality (greater than). Hence, by defining circuit mean as the total required slack plus running time divided by the number of arcs of a given circuit, the minimal cycletime equals the maximum circuit mean:

$$\lambda^* = \max_c \left\{ \frac{\sum_{P_{ij} \in c} (m_{ij} + \sum_{(k,j) \in P_{ij}} a_{kl})}{|c|} \right\}.$$ \hspace{1cm} (15)

This characterisation can be used to construct more efficient algorithms for finding the minimal cycletime other than using LP solvers (see appendix B).
4 Conclusions and further research

In this paper we formulated the problem of finding a timetable with optimal buffer times that is both fast as robust with respect to delays. Given buffer times and distributions of initial delays, we obtain delay propagation and waiting time distributions. The other way around is more difficult: We do not have an elegant way to translate constraints on waiting times into constraints on delay propagation other than the "trial and error" approach. However, when constraints on delay propagation are given, the problem of finding optimal buffer times can be formulated as a linear programming problem that can be solved by standard optimisation techniques. Moreover, for cyclic networks the minimal cycle time (corresponding to the optimal buffer times) is characterised as the maximal circuit mean, giving rise to more efficient algorithms.

Further research should investigate how to find appropriate constraints on the delay propagation and how to extend the analysis to stochastic running times.

Acknowledgements

The author would like to thank Erling D. Anderson and Kees Roos (both members of the Optimisation Group at TU Delft) for some useful discussions.
References


Appendix A: A circuit in the active lp-constraints

**Theorem 1:** The active constraints corresponding to a (finite) solution of problem (12)-(13) always contain a circuit.

**Proof:** Consider a solution of problem (12)-(13) and the set $c^*$ of active constraints. Without loss of generality we may choose at most one constraint for each pair $(i,j)$ and we denote the corresponding paths by $p_{ij}$. Consider the subproblem:

$$
\begin{align*}
\min & \quad \lambda \\
\text{s.t.} & \quad v_i - v_j + \sum_{(k,l) \in p_{ij}} \lambda \geq b_{ij} \quad \forall p_{ij} \in c^*
\end{align*}
$$

where $b_{ij} = m_{ij} + \sum_{(k,l) \in p_{ij}} a_{ij}$. The dual problem is formulated as follows (cf. [5]):

$$
\begin{align*}
\max & \quad \sum y_{ij} b_{ij} \\
\text{s.t.} & \quad \sum y_{ij} = \sum y_{ji} \quad \forall i \\
& \quad \sum |p_{ij}| y_{ij} = 1 \\
& \quad y_{ij} \geq 0 \quad \forall i, j
\end{align*}
$$

This problem is known as a circuit flow problem, except that we have another normalisation constraint ($\sum_{ij} |p_{ij}| y_{ij} = 1$ instead of $\sum_{ij} y_{ij} = 1$). It is well known that a circuit flow problem is feasible if and only if the underlying network contains a (directed) circuit (cf. [4]). So, the active constraints must contain a circuit since we assumed the original primal problem to be bounded ($\lambda > -\infty$).
Appendix B: The Floyd-Warshall algorithm

The well known Floyd-Warshall algorithm (cf. [1], [5]) can be adapted for computing the minimal cycletime characterised by (15):

First, we observe that when subtracting a constant \( \mu \) from all running times, the minimal cycletime decreases by the same constant:

\[
\max_c \left\{ \sum_{\rho \in C} (m_{ij} + \sum_{(k,l) \in \rho} (a_{kl} - \mu)) \right\} = \max_c \left\{ \sum_{\rho \in C} (m_{ij} + \sum_{(k,l) \in \rho} a_{kl}) - |C| \mu \right\}
\]

\[
= \max_c \left\{ \sum_{\rho \in C} (m_{ij} + \sum_{(k,l) \in \rho} a_{kl}) \right\} - \mu = \lambda - \mu
\]

Let \( \mu \) be a lower bound of the minimal cycletime: \( \lambda \geq \mu \). Next, we check whether the network with the adapted running times contains a positive circuit. If not, we have \( \lambda \leq \mu \) by (B1) and we conclude that \( \mu \) equals the cycletime \( \lambda \). If the adapted network does contain a positive circuit, we increase \( \mu \) by the corresponding circuit mean (\( \mu \) remains a lower bound of \( \lambda \)) and we repeat the procedure with this new \( \mu \) until we have found the minimal cycletime.

So, the remaining question is how to check the existence of a positive circuit given a constant \( \mu \). For this purpose we adapt the Floyd-Warshall algorithm:

**Adapted Floyd-Warshall algorithm:**

Compute the following matrices:

\( A_0 := A - \mu \),
\( S_0 := A_0 + m \),
\( A_k : \{A_k(i,j) := \max\{A_{k-1}(i,j), A_{k-1}(i,k) + A_{k-1}(k,j)\}, k = 1,2,...,n\} \),
\( S_k : \{S_k(i,j) := \max\{S_{k-1}(i,j), S_{k-1}(i,k) + S_{k-1}(k,j), A_{k}(i,j) + m\}, k = 1,2,...,n\} \),

and decide the existence of a positive circuit:

\( \text{positive} := (\max\{S_n(i,i) > 0\}) \).

Here, we may interpret \( A_k(i,j) \) as the maximal (adapted) running time of a path from \( j \) to \( i \) via nodes in \( (1,...,k) \) and \( S_k(i,j) \) as the maximal (adapted) running time plus required slack of a path from \( j \) to \( i \) via nodes in \( (1,...,k) \).

The algorithm suggests that the adapted network contains a positive circuit if and only if at least one diagonal element of \( S_n \) is positive. We refer to [1] for the proof of this fact and for the computational complexity of this algorithm.
Appendix A: Additional Results and Proofs

Theorem 1: Given a graph $G = (V, E)$ with non-negative weights on its edges, we can construct an optimal solution to the minimum cost flow problem in time $O(mn^2)$, where $m$ is the number of edges and $n$ is the number of nodes.

Proof: The proof follows a similar approach to the one used in the original paper. We construct a series of linear programming problems, each of which is solved using the Simplex method. The key insight is that if we start with an empty flow and iteratively add flows that are minimum cost with respect to the current residual network, then we will eventually obtain an optimal solution.

Let $a_i$ denote the number of flows assigned to edge $i$. Then the dual problem is:

$$\min \sum_{i} b_i a_i$$

subject to:

$$\sum_{i} c_{ij} a_i \geq b_j$$ for all $j \in V$.

The optimal solution to this problem gives the optimal flow. Each iteration of the algorithm selects an edge $i$ for which $c_{ij} > 0$ and $b_j$ is maximized subject to the constraints.

The algorithm terminates when the residual network is empty or when no further improvements can be made.

Corollary: The minimum cost flow problem can be solved in polynomial time for a network with $n$ nodes and $m$ edges.

Proof: By Theorem 1, the maximum flow problem can be solved in $O(mn^2)$ time. Since the minimum cost flow problem can be reduced to the maximum flow problem, it can also be solved in polynomial time.

The algorithm above is a dual ascent method. However, there are other algorithms that can solve the minimum cost flow problem in different ways, such as the Ford-Fulkerson algorithm and the Dijkstra's algorithm.

For more information, please refer to the original paper for a detailed explanation and proofs.
Accessibility service levels as an instrument to improve the accessibility of special activity areas

TRAIL Research School, Delft, December 1999

Author
Ir. A. van der Elst
Faculty of Technology, Policy and Management, Delft University of Technology

Thesis supervisor
Prof.dr.ir. R.E.C.M. van der Heijden
Faculty of Technology, Policy and Management, Delft University of Technology
Abstract

The growing scarcity of accessibility has been reported in several studies, indicating that many urban areas are becoming inaccessible to people with mobility issues. This trend is exacerbated by the increasing demand for public transport services, which often results in overcrowded and inefficient systems. The problem is further compounded by the increasing number of private vehicles on the road, leading to congestion and reduced accessibility. The emphasis on sustainability and environmental considerations has also led to a push for more efficient and accessible public transport systems. However, the challenge lies in balancing the needs of different stakeholders, including private operators, government, and the public. This paper aims to provide an overview of the current state of accessibility in public transport systems and to explore potential solutions to improve accessibility and reduce congestion. The case study of Utrecht City Project is used as an example to illustrate the application of these solutions.
Contents

Introduction 4

The Growth of Security at Societies of Peace 5

Private Intelligence Various Houses of Intelligencesrc

Assessment-Service-Fenale 4

Case Utilinity City Project 8

Conclusion 10

References 12
Abstract

In the Netherlands, for long the well-known problems related to growing mobility (traffic congestion, use of scarce space, environmental deterioration, decreasing access to economic centres) have been regarded as a responsibility of public authorities. Both entrepreneurs and drivers tended to consider themselves victims, yielding an attitude of looking at local, regional and national authorities to solve these problems. Strategies for solution have long been dominated by building car-focused infrastructures (roads, parking facilities). Non-infrastructure strategies (such as e.g. selective pricing, serious behavioural changes, significantly better public transport services, large scale business related transport management, or modal shift towards rail and water) have long received less attention.

The new mind-set in the transport policy also effects the attitude of entrepreneurs with shopping, recreational or conferencing facilities regarding issues such as locational decision making, accessibility for visitors and use of alternative transport modes. Many of these facilities are located at spots difficult to be reached by car, for instance in inner cities. The problem with several locations for instance is the enormous amount of cars to be parked at top-days and due to that congestion at local roads, time lost for parking entrance and exit and unattractive walking distances to the location. Hence, customers experience growing difficulty in reaching these locations. Increasingly, entrepreneurs are aware of the fact that they have to take initiative themselves in order to offer visitors attractive transport alternatives to reach the destination. The meaning of 'place' is changing from 'being there' to 'being easy to reach'. Access to the location is becoming a part of the marketing mix for the activities employed at that location.

Accessibility will proof to be a factor of great importance in the competitiveness of locations. Until now mobility has been seen as a collective good. The government therefore dictated the nature of the service and the terms (i.e. pricing). Now that accessibility is becoming scarce, the traditional role of the government is no longer sufficient to guarantee a sufficient accessibility for all locations. As a result the interest of private parties in accessibility is growing. Nevertheless, private parties are only willing to invest in accessibility if this has a direct influence on the size of their target group and turnover.

The private parties' involvement asks for a restructuring of the mobility market from supply focused to demand focused. But little knowledge exists of the properties of the demand side on the mobility market. Before private parties are willing to contribute to a better accessibility more knowledge about the effects of new solutions to improve accessibility and a new approach to support the decisions for investment are necessary. Because of this a Phd-survey is started with the following question being the main-point of attention.
"To which requirements do mobility-services have to come up to not only to improve the accessibility of the location but also to make these services attractive for the customer and the location-owner?"

The concept of accessibility-service-levels, is a way to find these requirements and to support decisions of private parties about whether or not to invest in accessibility. The concept is discussed further in this paper and the Utrecht City project is used as a case to develop the concept and to get some experience with it.
1. Introduction

In the Netherlands, for long the well-known problems related to growing mobility (traffic congestion, use of scarce space, environmental deterioration, decreasing access to economic centres) have been regarded as a responsibility of public authorities. Both entrepreneurs and drivers tended to consider themselves victims, yielding an attitude of looking at local, regional and national authorities to solve these problems. Strategies for solution have long been dominated by building car-focused infrastructures (roads, parking facilities). Non-infrastructure strategies (such as e.g. selective pricing, serious behavioural changes, significantly better public transport services, large scale business related transport management, or modal shift towards rail and water) have long received less attention.

Nevertheless, since the second half of the nineties one can experience society-broad a growing awareness of the problem complexity. Incentives to this mental shift have been given by several institutional changes, reducing the traditionally important role of public authorities and strengthening private initiatives. This is in particular clear for the sector of public transport, but also in the field of infrastructure building, maintenance and exploitation serious changes are introduced (van der Heijden et al 1999).

The new mind-set in the transport policy also effects the attitude of entrepreneurs with shopping, recreational or conferencing facilities regarding issues such as locational decision making, accessibility for visitors and use of alternative transport modes. Many of these facilities are located at spots difficult to be reached by car. The problem with several locations is for instance the enormous amount of cars to be parked and due to that congestion at local roads, time lost for parking entrance and exit and unattractive walking distances to the location. Examples of locations that have done studies for solutions to the parking problem are the conferencing centre “De Jaarbeurs” in Utrecht (Arends en Samhoud 1991), the entertainment parc “Dolfinarium” in Harderwijk (Versnel 1998) and the headoffice of the “ANWB” in Scheveningen (Bos 1999).

Governments as well as private parties gain from minimising or combating congestion and maximising accessibility. As mobility grows, congestion increases further and the accessibility of cities and economic centres will become more and more problematic. A new development is to think in terms of commerce to consider accessibility as a way to achieve customer’s loyalty. Accessibility is attractive to private parties when -and in the extent to which- they help to strengthen the commercial attractiveness of the locations concerned. From this perspective accessibility is a part of the product of the location-owners. Entrepreneurism can play a role by developing additional accessibility quality.
The private parties' involvement asks for a restructuring of the mobility market from a supply-focused to a demand-focused but little knowledge exists of the properties of the demand side on the mobility market. Before private parties are willing to contribute to a better accessibility more knowledge about the effects of new solutions to improve accessibility and a new approach to support the decisions for investment are necessary. Because of this a Phd-survey is started with the following question being the main-point of attention:

"To which requirements do mobility-services have to come up to not only to improve the accessibility of the location but also to make these services attractive for the customer and the location-owner?"

An important instrument in the orientation to the consumer's demands concerned to accessibility is the concept of accessibility-service-levels. An accessibility-service-level expresses what the consumer thinks to be acceptable for reaching his destination from his origin. When the accessibility service levels are known, it is easier to find the solution that has the biggest effect on the quality of accessibility and the behaviour of the customer.

Thinking in terms of accessibility-service-levels makes it possible to integrate the interests of the location-owners with the additional quality of accessibility. The interested location owner can offer extra quality on top of the basic provisions for which he pays himself or charges the traveller or shopkeeper. The location-owner has the ability, in co-operation with the government, to take control of the finances and exploitation of venerable transportation, which doesn't have to imply extra infrastructure. This way the concept of accessibility-service-levels can help to decide which accessibility solution is best.

This paper elaborates the concept of accessibility-service-levels. Chapter 2 describes how the growing scarcity leads to involvement of private parties. After that chapter 3 shows why different points of view, from both public and private parties, lead to problems with public-private co-operations concerned to accessibility. Chapter 4 describes the idea of accessibility-service-levels as a concept that can support decision-making on accessibility. Points of attentions are the way in which accessibility-service-levels can contribute to public private negotiation-processes and to an improvement of the results. In chapter 5 the first experiences with accessibility-service-levels in the Utrecht City project (UCP) are described and in chapter 6 at least, some conclusions are drawn.

In the Phd-survey that is bounded with this paper the UCP is taken as a case. In Utrecht accessibility is one of the subjects that plays an important role in the public private partnership for the UCP. Public and private parties have to make deals concerning accessibility and therefore they would like to know more about the concept of accessibility service levels. This gives an excellent opportunity to test the concept and to do some experience with it.
2. The growing scarcity of accessibility

With 15 million inhabitants, the Netherlands belongs to the most highly populated areas on earth. In the next 20 years this number will increase two more million. More and more locations in the Netherlands will be difficult to reach. The government has made some plans to guarantee the accessibility of locations in the Netherlands but is seems that the goals are difficult to reach.

The perifere roads are overloaded and public transport has not the quality to be a good alternative to the car. Guaranteeing a good accessibility is a growing problem of more and more economic centres in the Netherlands. Some parties that cope with problems concerning accessibility are: the business-area in the south east of Amsterdam (DynaVision 1999), the partners of the Utrecht City Project (van der Elst 1999b), the entertainment parcs “Efteling” and “Dolfinarium” and even some business-area in the south of the Netherlands (TC&O 1999).

It is of preliminary importance for these locations that people are able to visit the locations. When the locations cannot be reached they can’t exist. The law of BREVER says that people spend a fixed amount of time to travelling per day (Hupkes 1977). If they have to spend more time per trip because of congestion they will make less trips. Besides that for some activities people want to spend of maximum amount of time on travelling. If the amount of time exceed this level people will look for an activity on a location that is easier to reach. In table 2.1 is shown how many hours are spend on congestion on the mainroads in the Netherlands in the last years (AVV 1998). The problem is still growing and more and more people will decide to stay at home in stead of losing a lot of time on the roads on their way to visit a location.

| Table 2.1 Number of hours lost by congestion on the mainroads (* 1 million) |
|-----------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Persons (cars)                   | 13,4        | 13,7        | 16,2        | 17,0        | 17,3        | 17,9        |
| Goods (cars)                     | 1,1         | 1,2         | 1,4         | 1,5         | 1,5         | 1,6         |
| Total number of cars             | 11,2        | 14,5        | 14,9        | 17,7        | 18,5        | 18,8        | 19,4        |

Traditionally guaranteeing a certain quality of accessibility was mainly regarded as a responsibility of public authorities. Last years the government has taken a lot of initiatives with financial help for market-parties and a lot of notes, plans and surveys. Some examples are the Dutch Second Transport Structure Plan in 1990, the note: about working together on accessibility (“samenwerken aan bereikbaarheid”), the note about seamless multimodal mobility (“Ketenmobiliteit en Dienstverlening”), and so on. With the growing scarcity of accessibility, accessibility becomes more and more important, not only for public but also for private parties (van Egeraat 1998).
Each year Dolfinarium Harderwijk draws 1 million visitors and it wants to expand to 1.6 million in 2005. 80 percent of all the visitors comes by car. A good accessibility, nurtured by an optimal traffic-flow and adequate parking-facilities is therefore a prime requisite for the future of the Dolfinarium. The available parking-spaces are already insufficient at this moment. In the last few years the Dolfinarium itself has developed a number of concepts to increase the accessibility. The Dolfinarium has been stimulating an investigation of the possibility for over 6 years now for realising a large-scale parking-facility outside of the centre. Despite the willingness of the Dolfinarium to co-finance no agreement has been reached with the municipality. A solution is still hard to find.

The workshops: "Ketenmobiliteit" (held during the Intertraffic March of 1998) show that more and more organisations see accessibility as an essential aspect of service providing to the customer (for which one is willing to pay). This mainly concerns the so-called "location-owners": shopkeeper-organisations, recreational businesses, congress-centres and real estate companies. Next to optimising their core-activities these organisations pay attention to accessibility as a partial product.

Because of the involvement of private parties a new group of players is being created on the mobility market. The parties that traditionally are involved with the decision-making concerning infrastructure are not (sufficiently) utilising these new developments and there possibilities are still there for the taking. An example of a location owner that is very concerned about his accessibility and has taken some initiatives is the "Dolfinarium Harderwijk".

Each year Dolfinarium Harderwijk draws 1 million visitors and it wants to expand to 1.6 million in 2005. 80 percent of all the visitors comes by car. A good accessibility, nurtured by an optimal traffic-flow and adequate parking-facilities is therefore a prime requisite for the future of the Dolfinarium. The available parking-spaces are already insufficient at this moment. In the last few years the Dolfinarium itself has developed a number of concepts to increase the accessibility. The Dolfinarium has been stimulating an investigation of the possibility for over 6 years now for realising a large-scale parking-facility outside of the centre. Despite the willingness of the Dolfinarium to co-finance no agreement has been reached with the municipality. A solution is still hard to find.

As a second important initiative the Dolfinarium has created a transportation-plan in co-operation with NS -Reizigers and MIDnetgroep. In this plan a diversity of ideas is being explored and developed focused on collective transport. Collective transportation in the recreational sector is hard to organise because of the rugged patterns of transport in space and time. People from all over the countries come to visit the Dolfinarium, and these flows of visitors are also affected the holidays, vacations and the weather. Despite these handicaps the plan of transportation has concrete solutions with a car-mobility reduction of 8%. A year after the proposition of the plan to the municipality most of the solutions have become outdated, due to amongst others different priorities with the providers of the collective transportation. All the delays that were created are and have been extremely damaging for a further healthy development of the Dolfinarium. This makes strategic reconsideration inevitable, unless new directions can be needed to create an optimal accessibility of the Dolfinarium in the future. And maybe one of these directions will be seamless multimodal mobility.

\[1\] The public transport operator in and for Utrecht and surroundings
The Efteling, another big theme park in the Netherlands, is also worried about accessibility. It is known people maximum want to spend one and half an hour for leisure-activities (van der Zijl 1999). In effect of the ever-growing congestion on the mainroad-network, the area of service of the Efteling is becoming smaller all the time. The board of directors wants to change this. Due to the lack of improvements in the public transport there still are not good alternatives for car-use. On its own initiative the Efteling has in co-operation with coachnet (service provider of the VSN-companies) decided to have a fairytale-coach going back and forth. These are special busses that drive to the Efteling from certain different places in the Netherlands.

All those initiatives indicate that the number of private parties that is interested in improving accessibility is growing. The first step is taken but it is not clear how private parties can really contribute to a better accessibility and what the best solutions (investments) are. For private parties working together with public parties is often difficult and to slow, also because the diverging interests and goals (van der Heijden et al 1999). In the following chapter the public and private interests and the differences will be further discussed.
3. Private interests versus public interests

The Dutch Second Transport Structure Plan, in 1990 accepted as the basis for public transport policy development, recognised the relevancy of broad strategies based on the simultaneous application of various instruments. The plan, compared to public policy plans in other Western countries (Banister 1994 xiii), offers a new perspective on the increasing problems of the transport system and the ways to cope with them. Much attention is paid to increasing accessibility and mobility without decreasing liveability. These three items are not always easy to combine. Increasing accessibility by building new infrastructure results in more car-use and doesn’t have a positive effect on the liveability. That’s why solutions are often sought in improving public transport and shifting the modal split to less car-use.

For private organisations (the location-owners) the customer usually is the critical factor. Location-owners want to reckon their service-area and their number of visitors. The question for the location-owners concerns the levels of accessibility that is desired and the factors (of accessibility) that do influence the choice of the customer (to go/ not to go/ going to a rival). For the location-owners it’s of great importance that people visit the specific location. Of course not only accessibility is an important factor for the customer, but a number of factors are taken into consideration. The decision of a customer whether or not to visit a location is determined by a couple of factors related to the attractiveness of a location, like the diverse activities on locations, the liveableness and the safety. Accessibility is also one of them (figure 3.1)

![Accessibility Service Levels](image)

*Figure 3.1 The attractiveness of locations*
The attractiveness of a location implies the value that a consumer gives to a location. This value will differ per customer-segment. Examples of factors of attractiveness are assortment, safety, activities on location and accessibility (Stemerding 1996). Accessibility itself consists of more factors too. The accessibility of a location is defined by "the afford that a consumer has to take to reach the location". Accessibility is determined by a.o. travelling-time, travelling-costs and comfort. Accessibility is related to all modalities including mobility-chains (van der Elst 1998). It is interesting to know more about the value of each accessibility- and attractiveness-factor per customer-segment and to know what the room is for improvement. In Utrecht, public and private parties decided the want to know more about these facts and they started a survey. Some results of this survey will be discussed later.

In summary, the interests of public organisations focus mainly on the traffic flow and on minimising congestion. Solutions are not only sought by building new roads but also by offering new modes of transport like chain-mobility and by improving the existing public transport. Private parties on the other hand focus mainly on the behaviour of customers. They want to stimulate mobility and attract as much customers as possible; therefore they need a good accessibility, not only by public transport but also by car.

Both, public and private parties, in a way focus on the behaviour of people. Public parties want people to use new travelling-concepts and new modalities instead of the car; private parties simply want people to reach them. Both parties want to find new mobility concepts that meet the customers' demands. Both want the customer to use new concepts (other than - or in combination with - the car). To know more about the customer, his behaviour and his demands, the concept of accessibility service levels can be of great help.
4. Accessibility-service-levels

Accessibility-service-levels are indicators for the levels of the different components of accessibility that people still consider as acceptable to reach a certain destination. An accessibility-service-level can be expressed in different components like travelling-time, but also in components like price, comfort and ease of use. The accessibility service level is determined by the customer himself: what the customer thinks is acceptable depends on what he thinks is reasonable and on the effort it takes to reach a competing destination.

The accessibility-gap is the difference between the real accessibility and the accessibility-service-level. On this moment chance of congestion is mainly used as an indicator for the decisions about investments by public authorities. In the "Randstad" the chance on congestion is allowed to be 5% and on the connections with the hinterland this is 2% (Ministry van V&W 1997[1]). If this number exceeded, then new infrastructure should have to be created. The danger with this level is, thought thoroughly through, that the only solutions are searched in the creation of extra road-infrastructure.

The target of accessibility-service-levels is that the discussion between public and private parties about accessibility will be supported by the knowledge about the demands of the customers. Both, public and private parties want to find solutions that are used by the customers. To find these solutions they need more knowledge about the behaviour of those customers. Therefore the concept of accessibility-service-levels can be of great help. In figure 4.1 is shown when public and private parties seek each other to think about accessibility and when accessibility service levels are important. This figure is developed for Utrecht City Project[11].

Initially the government is responsible for the accessibility of locations. Public authorities invest in infrastructure, public transport and new technologies to bring accessibility to an acceptable level (step 1 figure 4.1). If after these investments the accessibility of a location still is not sufficient to private parties (to draw enough customers), private parties are (initially) willing to contribute to an improvement of the accessibility of locations (step 2, figure 4.1). That is when public and private parties are driven together.

Private parties utilise another acceptable level of accessibility than public parties. Do public parties mainly focus on the level of congestion, private party's look more at the amount of accessibility (and the accompanying price tag) that the location needs to draw (sufficient) customers. Though even when the accessibility is not well enough compared to the criteria, private parties only invest in accessibility when it has a direct influence on the mobility-choice of their target-group and leads to an increasing profit. In fact only a few steps have to be followed to determine whether it is interesting (for private parties) to invest in accessibility or not (figure 4.1).

---

Accessibility Service Levels
In step 2 the complete attractiveness of a location is compared to that of the competition. Only if the competition scores higher something has to be done. In step 3 is determined what the share of accessibility is in the attractiveness-mix and what the size of the accessibility-part and the room for improvement are. This deals with the question whether accessibility is the bottleneck in the attractiveness of the location, i.o.w. is the accessibility-quality so bad, that the amount of visitors will increase when it is improved? If the answer to this is confirming than the accessibility-gap is determined in step 4.

For every component of accessibility the accessibility-service-level of the consumer and the current value will be determined. This way a view of possible options of improvement is created. As a final step solutions and possibilities are sought to close the gap and whether this will be profitable. Step 4 is related to the comparison of the accessibility with the accessibility service-level that is determined by the customers themselves. There still is less knowledge about the demands of the customers; there is a great difference between what the stakeholder thinks that the customer wants and what the customer actually wants (van der Elst 1999\textsuperscript{9}). The stakeholder will have to gather more knowledge about the customer. In the final (fifth) step the costs of the options to close the accessibility-gap are determined and measured against the benefits.

\textbf{Figure 4.1 Public and private parties working together on accessibility}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{accessibility_diagram.png}
\caption{Public and private parties working together on accessibility}
\end{figure}
5. Case Utrecht City Project

The Utrecht City Project (UCP) is a plan for reconfiguration of the area around Utrecht central railway station. The present number of shops, cafes, offices, conference facilities and apartments will substantially be increased. The Utrecht central railway station is one of the largest nodes in the Dutch railway network. The station and its facilities have to be adapted to new developments, such as changes in the present rail services and the introduction of high-speed trains. UCP intends to enlarge the regional and national meaning of the area, resulting in about 75 million visitors to the Utrecht central railway station yearly. It is clear that tackling the transport and accessibility problems of this scale is a major challenge.

In order to cope with this challenge, a Platform UCP has been established. Key players, among who the owner of the shopping mall Hoog Catharijne (WBN), different parts of the Railway Company (NS), the city authorities and the owner of a large conference centre (Jaarbeurs), participate in the Platform. The Platform is the bases for public-private partnership in project development. Part of the public private partnership is an arrangement concerned the accessibility of Utrecht.

It seems difficult to achieve agreement on accessibility because of the different interests of the parties. The private parties put more emphasis on the car-accessibility, whereas the public authorities want to stimulate the access by high quality public transport. The private parties want the public authorities to guarantee a certain level of car-accessibility. The public parties on the other hand can't give any guarantee and besides that there is discussion about the level of accessibility that is needed. So, the question is relevant what level of accessibility is demanded for different market segments and which solutions likely are to improve the accessibility.

To find out more about the demands of the customers concerned accessibility an investigation has started to the accessibility-service levels for different groups of customers. Therefore the customers of UCP have been shifted in 10 groups. These groups are fixed from a marketing-perspective, not from a traffic point of view (van Egeraat 1998). This is done because the customer is most important and new mobility services have to fit the different groups of customers. To split the customers a number of criteria have been taken into consideration: the alternatives one has to go to another location, the way one is forced to travel on a certain time, the time some one stays at a location and the motive for making the trip. This has led to the customer-segments named in table 5.1.
Table 5.1 Customer-segments in UCP

| 1. | People that go shopping for fun          |
|    | People that go shopping to buy non-food |
| 3. | People that go shopping to buy food     |
| 4. | People that need to go to the Utrecht Central Rail Station to travel by public transport |
| 6. | People that have a business meeting in UCP |
| 7. | People that go to a conference or a congress |
| 8. | People that go to a event or a change during the day |
| 9. | People that go to a musical or other entertainment during the evening |
| 10. | People who live in UCP |

The following step is to investigate the accessibility-service-levels for these segments. Therefore it is necessary to identify the most important accessibility-components per segment and specify these accessibility-components so that they can be measured. In view of the project accessibility-service-levels for the Utrecht Centre Project (UCP) already little investigative research has been made in accessibility and accessibility-service-levels in Utrecht (van der Elst 1998). Two important conclusions of this research were:

1. That the stakeholders who are involved (Jaarbeurs, Winkelbeleggingen Nederland, Gemeente Utrecht, Rijkswaterstaat Utrecht and NS) are not well informed about the demands the customer sets about accessibility. For many aspects the stakeholders had more stringent norms than the customer himself did. An example of this is the waiting time during a transfer, shown in figure 5.1.

2. For some components of accessibility, there was found an accessibility-gap. The maximum level the stakeholders had set was lower than the real value. This is also true in the case of waiting time during a transfer.

![Figure 5.1 The gap of the waiting-time during a transfer.](image)
Remarkable is that the levels of all stakeholders are lower than the actual waiting-time and also lower than the level the customer himself has set. The "Jaarbeurs" is the only party that has a good impression of the maximum level of waiting-time of the customer. This picture shows that there is a gap between the current waiting-time and maximum acceptable waiting-time during a transfer for both the levels of the customers and the levels of the stakeholder.

Another example concerns the information-provision during a trip. Figure 5.2 shows that for car-users only 8% is not satisfied with the quality and quantity of information during the trip. For public transport this is more than 35%. This shows that there is lot of improvement possible. The figure also shows that the actors accept a bigger percentage of unsatisfied customers for public transport than for car-use. This is striking because car-use is private transport and a car-user should take care of his own information, public transport on the other hand is a product that should contain the needed information.

The next step in the survey for the Utrecht City Project is to set up a bigger survey (with about 3000 respondents) to find the accessibility service levels for all customer segments for the most important factors of accessibility. Therefore of course we have to find out first which factors of accessibility are most important; this will differ per customer-segment.
6. Conclusions

Location owners are increasingly aware of the problems of local access by car and fear a lower attractiveness and a decreasing market share. Therefore, they are more and more willing to take responsibility and help to realise innovative concepts for accessibility. In terms of marketing, these innovative concepts can be considered as a part of the marketing strategy of the location owners. Location owners though don't know exactly what level of accessibility the customer demands and how important accessibility is as a part of attractiveness. This paper addressed the useful help the concept of accessibility-service-levels can offer to find solutions to improve the accessibility of special activity areas. In the concept of accessibility-service-levels the customer and the way he or she makes a choice between competing destinations, is the starting-point. In the process in which a customer decides whether or not to visit a location he weighs the attractiveness of possible destinations against the afford to reach these destinations, i.e. the accessibility. The different components of accessibility have different values to the customer and by changing these components the accessibility can be made better or worse. When the maximum accepted levels of accessibility-components are known these can be used to support the choice for an investment or implementation of a solution.

In the PHD-survey that is bounded to this paper, there will be further investigation on the relation between accessibility and attractivity and on the components of accessibility. The following question will be the main point of attention.

"To which requirements do mobility-services have to come up to not only to improve the accessibility of the location but also to make these services attractive for the customer and the location-owner?"

One of the following steps now is to do theoretical research to accessibility and to the experience with new mobility-services. Another step is to do a big research among customers to the different components of accessibility and their values. If more is known about the different components of accessibility and their values per customer segment some mobility-service can be developed and tested in cases.
References


v DynaVision Management Consultancy (1999): Het verbeteren van de bereikbaarheid van het bedrijvencomplex Amsterdam Zuid Oost, Oktober 1999 (in Dutch)

vi TC&O Management voor mobiliteit (1999), Ketenmobiliteit voor bedrijfsvervoer in Limburg, Eersel, April 1999 (in Dutch)


xii Zijl, R.R.H.M. van der (1999), Ketenvorming begint bij versterking schakels, in: congresmap Ketens in Beweging, Nederland Haarlem, April 1999 (in Dutch)


References


Improving Punctuality and Transfer Reliability by Railway Timetable Optimization

TRAIL Research School, Delft, December 1999

Author:
Drs. Rob M.P. Goverde
Transportation Planning and Traffic Engineering Section,
Faculty of Civil Engineering and Geo Sciences, Delft University of Technology

Thesis supervisors:
Prof. Dr.-Ing. Ingo A. Hansen
Transportation Planning and Traffic Engineering Section,
Faculty of Civil Engineering and Geo Sciences, Delft University of Technology

Prof. Dr. Geert Jan Olsder
Department of Technical Mathematics,
Faculty of Information Technology and Systems, Delft University of Technology
Improving Productivity and
Transit Reliability of
Railway Timetable Optimization

TRAIL Research Group, DeL. November 1988

Abstract

The need for efficient transit systems has led to the development of new methods for timetable optimization. This paper presents a new approach to the problem of timetable optimization that incorporates the use of computer simulation techniques to evaluate the performance of different timetable proposals. The approach is based on the use of a genetic algorithm to search for the optimal timetable, and the results are presented in a format that allows for easy comparison with other proposals.

Introduction

The need for efficient transit systems has led to the development of new methods for timetable optimization. This paper presents a new approach to the problem of timetable optimization that incorporates the use of computer simulation techniques to evaluate the performance of different timetable proposals. The approach is based on the use of a genetic algorithm to search for the optimal timetable, and the results are presented in a format that allows for easy comparison with other proposals.

Results

The results of the simulation experiments are presented in Table 1. The table shows the performance of the proposed timetable compared to the current timetable. The results indicate that the proposed timetable is superior in terms of both passenger satisfaction and operational efficiency.

Conclusion

The results of this study suggest that the proposed approach to timetable optimization is effective in improving the performance of transit systems. Further research is needed to evaluate the robustness of the approach and to develop a practical implementation strategy.

References


## Contents

**Abstract**

1 **Introduction** .................................................. 1

2 **Modelling Timetable Constraints** .......................... 4
   2.1 Introduction .................................................. 4
   2.2 Time Window Constraints .................................. 5
   2.3 Cycle Constraints ............................................. 10
   2.4 Rolling Stock Constraints .................................. 12

3 **Railway Timetable Optimization** ......................... 16
   3.1 Introduction .................................................. 16
   3.2 The Periodic Railway Timetable Optimization Problem .... 16
   3.3 Computational Complexity .................................. 17
   3.4 Preprocessing: Integer Bounds .............................. 18

4 **The Network Synchronization Problem** ................... 20
   4.1 The Optimization Problem .................................. 20
   4.2 Tight Connections ............................................. 21
   4.3 Reliable Connections ......................................... 22
   4.4 Transfer Connections .......................................... 24

5 **Conclusions** .................................................. 28

**Acknowledgements** ............................................. 28

**References** ....................................................... 28
Abstract

The Dutch railway network is operated close to capacity with the current safety system. This leaves little space for control by process operators by which a delayed train can cause severe delay propagation. The NS currently invest a large amount in punctuality improvement of the railway operations. This paper shows how the effect of delays can already be incorporated in the timetable design process to obtain robust and optimized timetables for a punctual and reliable operation.

A detailed modelling approach is presented to compute periodic network timetables with an optimal distribution of buffer times at those points where they are needed the most without being too conservative. The variables are the buffer times and the proposed objective is to minimize a weighted sum of individual costs for each buffer time. Here, the weights define the relative importance of the buffer time costs. The following cost functions are proposed and proved to be convex for general running time distributions: buffer time (fast travel times for through passengers and small operating costs); expected delay due to buffer time failure of rolling-stock and crew connections (increasing punctuality) and train circulations (stability of train circulations); and expected transfer waiting time (trade-off between small transfer times and transfer reliability).

The constraints on a feasible timetable are modelled conveniently by time window constraints concerning interactions between two periodic events (arrivals and departures). This is compatible with the timetable design system DONS of the NS. For the optimization problem these constraints are reformulated using the buffer times as variables rather than the periodic event times. The resulting cycle constraints state that the oriented sum of process and buffer times on fundamental cycles of the constraint graph is a multiple of the cycle time. The buffer times are not periodic by which they can be used straightforwardly in the objective functions. The periodic time windows are thus replaced by periodic cycles.

Theoretical relevance
The periodic timetable optimization problem extends the associated feasibility problem, the Periodic Event Scheduling Problem (PESP), and improves several problem issues. The problem is formulated as an optimization problem with convex cost function and buffer times as variables rather than the periodic events which is most common in literature. Solving the resulting large-scale convex mixed-integer programming problem requires a synthesis of various techniques from linear algebra, network flows, integer programming, and convex programming.

Societal relevance
An optimal robust timetable contributes to improve performance and attractiveness of the railway operation, increase service quality to passengers, and decrease costs to railway operators and passengers due to delays. Moreover, a punctual and reliable railway operation may attract more potential passengers who currently use other modes (cars). This reduces the congestion problem on the Dutch highways and the associated environmental pollution.
Abstract

The classical problems of constructing the best estimate of a continuous system of variables from a set of observations are considered. Methods are developed for constructing linear and nonlinear estimates of the unknown functions and their derivatives with the help of differentiable basis functions. The method of constructing multipliers of the condition equations of the estimate is used to solve the problem. The properties of the developed methods and their stability are investigated. The results of the investigations are applied to the development of algorithms and programs for the solution of the problems in question.
1 Introduction

Punctuality is a major concern to all transportation companies, especially to an underlying transportation network with lots of interdependent services and low service frequencies (up to 6 services per hour). The level of punctuality largely influences the reliability of connections, including crew, vehicle, and transfer connections. On the other hand scheduled connections are a main source for secondary (or knock-on) delays (for example if a connecting service waits for a delayed feeder service to secure a transfer connection) and hence on its turn affects punctuality. An optimal coordination or synchronization of connected services is thus important to improve punctuality whereas at the same time reliability of connections is obtained. Both service punctuality and connection reliability are main indicators for the performance and attractiveness of a transportation network. This paper concentrates on railway networks as they represent a particularly interesting case.

Railway networks have an additional difficulty due to the safety and signalling systems which increase the possibilities of conflicting movements. Moreover, the train service network of the Netherlands Railways (NS) is highly interconnected and is served near to its capacity (for the present safety and signalling system). Additionally, the railway timetable in the Netherlands is an integrated periodic timetable, i.e., all train services run in regular intervals. As a consequence, the railway planners are faced with an extreme hard and time-consuming task to design a feasible timetable that satisfies all constraints as imposed by for instance the line system, required connections and infrastructure limitations. Therefore, specific attention to the synchronization of connected services is currently very modest.

A feasible timetable is by definition realizable (conflict-free) for punctual operation. However, perturbations during operation result in variations of train running and dwell times. Perturbations are for instance varying train characteristics, varying driving behaviour, fluctuations in dwell times depending on the number of alighting and boarding passengers, and extreme weather conditions. A timetable has to be robust to small variations in running times which is achieved by adding running time margins to the minimum running times. This way the (scheduled) running times can compensate for most of the variations. The more severe perturbations result in exceeding the scheduled running times, which give primary delays. Moreover, once a train is delayed it may hinder other trains or propagate the delay to connecting trains, resulting in secondary delays. The secondary delays are indirectly the result of the robustness of the timetable design.

A robust timetable design aims at minimizing secondary delays during operation and thereby improving punctuality and reliability.

A railway timetable consists of the scheduled arrival and departure times of all train services at each station. Implicitly this also gives the dwell times for each line and the transfer times between lines at transfer stations where lines meet. From an operational point of view the timetable also fixes the connection times for rolling stock and/or crew connections, and headways between trains (at e.g.
conflict points). All these process times consist of two components: a necessary minimum process time (e.g., minimum dwell time for alighting and boarding; minimum transfer time for alighting, walking to the departure platform and boarding; and minimum arrival headway between two arriving trains at a station) and a possible buffer time or recovery time which improves robustness with respect to delays.

We are concerned with a periodic railway timetable. Obviously, this imposes severe restrictions on the timetable. In combination with the network interdependencies of all train services, it is no surprise that there is an amount of slack in a timetable necessary to fit all process times in the regular interval pattern. The resulting buffer times lead to larger process times which at first glance is annoying. However, buffer times may be very desirable to improve robustness of the processes with respect to variations in process times. The challenge is to distribute the buffer times over the service network such that they are advantageous during railway operation. This is the periodic railway timetable optimization problem.

The development of the advanced design system DONS (Designer of Network schedules) [12, 25] is a great step forward to computing a feasible timetable that satisfies all constraints, or if no such timetable exists, giving information about conflicting constraints. DONS also features a timetable optimization option that can be used after having computed a feasible timetable. The optimization has then be understood with respect to the computed train (sequence) orders. So the train orders remain fixed which obviously results in local optimal solutions, i.e., better (feasible) timetables might be obtained if train orders are allowed to be swapped. Moreover, more advanced cost functions have to be defined to (sub)optimize the timetable with respect to operation reliability. Another improvement would be to obtain (nontrivial) multiple feasible solutions (if there are any). Currently, the computations stop when a feasible solution has been found. The existence of multiple solutions with perhaps more convenient train orders then remains unknown. This paper gives a solution to the above problems and thereby contributes to the further development of the DONS features. The problem of computing multiple feasible solutions to the system of time window constraints is also the subject of a parallel research project [24].

The optimization of railway network timetables has had some attention in the literature. The problem is typically modelled as a (linear) mixed integer program (MIP). Weigand [28, 29] reports a graphical oriented modelling approach. The MIP-formulation is based on cycles in a graph representing all stop and transfer connections. Infrastructure constraints are not dealt with explicitly. Nachtigall [19, 20] formalizes this approach and proves that the problem is \( \mathcal{NP} \)-complete. An alternative MIP-formulation is based on constraints imposed by interrelated train pairs resulting in time window constraints [15, 16, 18]. Here also infrastructure constraints (minimum headway between a train pair) are incorporated. The corresponding feasibility problem has been formalized by Serafini and Ukovich [26] and is known as the Periodic Event Scheduling Problem (PESP). They proved that PESP is \( \mathcal{NP} \)-complete, and hence also the associ-
ated optimization problem is \( NP \)-complete. The PESP-formulation is also used in the DONS-system [22, 23, 25]. A related modelling approach is the schedule synchronization problem in which the line schedules (running and dwell times) are fixed. The problem is to determine the optimal departure times of all lines at their starting terminal station such that the costs of the resulting transfer times at stations where the lines meet are minimized. This optimization problem can be formulated as a binary integer program (BIP) [13] or an integer program (IP) [2]. However, infrastructure constraints can not be handled trivially by these formulations. Domschke [7] proved that the problem is \( NP \)-complete. Note that this problem can also be modelled as an MIP using the time window (or PESP) constraints as it is a special case for fixed dwell times. Then also infrastructure constraints can be incorporated.

Some literature is more or less devoted to reliability and performance measures of transportation services. Hallowell and Harker [10] studied expected delays due to interference at meet/pass points on a partially double-tracked railway line. Carey [3] and Ferreira and Higgins [8] consider the minimization of train running times and expected arrival delays at a single railway line including knock-on delays. Carey (and Kwieciński) [6, 5] and Higgins and Kozan [11] have generalized these models to networks of various lines. Carey [4] moreover studied the behavioural response of actual running times if the scheduled running time is increased to minimize arrival delays. Knoppers and Muller [14] studied the expected passenger transfer waiting time including the probability of missing a connection. Finally, Meng [17] developed a model to determine the required average dwell and transfer buffer times at a railway station for random (Gamma-distributed) inter-arrivals of trains, incorporating the number of trains and connections, the number of platform tracks, average secondary delays due to connections and conflicting train paths, and the risk of exceeding platform capacity.

The paper is organized as follows. Chapter 2 shows how to model the timetable constraints and get a constraint system of the buffer times. Chapter 3 defines the general periodic railway timetable optimization problem and considers its computational complexity. The network synchronization problem is addressed in Chapter 4 where various convenient cost functions for the individual buffer times are derived to obtain either tight connections, reliable connections, or optimal transfer connections, whatever is appropriate.
2 Modelling Timetable Constraints

2.1 Introduction

The train service network in the Netherlands is operated according to a regular interval or periodic timetable, which is also popular in train service networks in most densely populated countries in Europe, for long-distance railway networks as in Germany, and other public transportation systems. A periodic timetable is convenient for passengers but imposes severe restrictions on the timetable. All arrival and departure times must fit a regular interval pattern. Additional difficulties in finding a railway timetable are the network interdependencies and infrastructure limitations.

A (periodic) network timetable is a feasible solution to a (periodic) constraint system that specifies all interactions between the train services in the network. We assume that a line system is given in advance and that the process times are deterministic and given as well. Here a line system is defined as a set of lines, the train type of each line, and the connections between the lines at stations where the lines meet. A line is a sequence of train services on a particular route between two terminal stations, and the train type (for instance express train or local train) determines the intermediate served stations on the route. We define a timetable point as a node in the service network where interactions between trains are observed. Examples are transfer stations, level-crossings, and emerging sections. Then our input data contains

- train running times between timetable points including dwell times at intermediate stops;
- minimum dwell times at transfer stations;
- minimum turn-around times at terminal stations;
- minimum transfer times between train pairs at stations where lines meet;
- minimum crew and rolling stock connection times between train pairs at stations where a crew transfer or train coupling/decoupling is required;
- minimum headways between train pairs at timetable points.

Additionally maximum process times may be given to prevent too large dwell times, transfer times, etc. The crew schedules are usually planned after the timetable has been established. However, some crew connections may be known in advance when a timetable design is based on an existing timetable, with some adjustments due to for instance completed infrastructure projects. Typical crew connections are for example endpoints of lines. An extension of the planning process may include a feedback to the timetable design after crew schedules and rolling stock circulations have been obtained. Then slight adjustments of the
timetable design may increase the robustness of the crew connections. Note that a crew connection does not have to occur at each timetable period.

A timetable is expressed in arrival and departure times of all services at each station. In the sequel it is more convenient to refer to an event which can be both an arrival or a departure. Generally, an event is the start or end of an activity. So a train departure from a station is an event that activates a train run to the next station, and an arrival at a station is the end of a train run from a preceding station. A periodic event is an event that repeats at a regular interval with cycle time $T$. For instance, a departure in an hourly timetable repeats every $T = 60$ min. An event time is the occurrence time of the event, e.g. the departure time.

If running times between stations are assumed deterministic (for planning purposes) then it suffices to consider departure events only: an arrival time can be given as the sum of a departure time at a preceding station and the train running time between the stations. This is convenient if large service networks are considered as this gives a reduction to half the number of variables. The variable $\tau_i \pmod{T}$ denotes the (departure) event time of the periodic (departure) event $i$. Below we express all timetable constraints in departure events. We also somewhat sloppy refer to $i$ as both the departure and the activated train service.

### 2.2 Time Window Constraints

Connections at transfer stations impose synchronization constraints on the train services. Also the track layout and the safety and signalling system of a railway network imposes limitations on train movements. For instance, trains can not cross a level-crossing at the same time, and a safety distance has to be respected for two trains running on joint track sections.

A train service is defined as a train trip between two stations. All timetable constraints can then be defined as an interaction (e.g. a connection) between two train services. For example, train 1 must depart no sooner than 2 to 5 min after the arrival of train 2 to allow a transfer. Such a constraint is called a time window constraint. Note that an interaction has a ‘direction’, i.e., train 1 waits for train 2, but the reverse does not have to hold unless also an interaction is defined from train 2 to train 1. An interaction of two services occurs for example at a stop, a transfer, or a conflict point (mutual track section). Each interaction produces a minimum process time and a buffer time. So a buffer time can for instance be a dwell buffer time, a transfer buffer time, or a headway buffer time.

Time window constraints specify that the time difference between two periodic event times is restricted to a particular time window. Let $\tau_i$ and $\tau_j$ be two event times. Then a time window constraint is

$$\tau_j - \tau_i \equiv l_{ij} + r_{ij} \pmod{T} \quad \text{and} \quad r_{ij} \in [0, r_{ij}^{\text{max}}].$$

(1)

Here, $l_{ij}$ is the minimum time interval between event $i$ and $j$, and $r_{ij}$ is an additional nonnegative buffer time that is bounded from above by a given maximum
buffer time $r_{ij}^{\text{max}} \in [0, T]$. Equation (1) has to be evaluated modulo the cycle time $T$ which means that an integral multiple of $T$ can be subtracted where necessary. Without loss of generality we can write (1) as the MIP constraint

$$\tau_j - \tau_i = l_{ij} + r_{ij} + z_{ij}T$$

and $r_{ij} \in [0, r_{ij}^{\text{max}}]$ or

$$l_{ij} \leq \tau_j - \tau_i - z_{ij}T \leq l_{ij} + r_{ij}^{\text{max}}$$

(2)

where $z_{ij} \in \mathbb{Z}$.

All timetable constraints can be modelled conveniently as time window constraints, see also Schrijver and Steenbeek [25]. The constraints can be classified in synchronization constraints, infrastructure constraints, and additional market constraints, which are considered sequentially below.

**Synchronization constraints** specify that a train service $j$ has to wait for a train service $i$ from a preceding station to guarantee a connection. A connection can either be a stop or transfer, and we thus distinguish between **stop constraints** and **transfer constraints**. Let $t_r^i$ be the running time of train service $i$ from the preceding station, and $t_{ij}^{\text{min}}$ be the minimum dwell/transfer time to the connecting train $j$. Then $l_{ij} = t_r^i + t_{ij}^{\text{min}}$ is the minimum time after the departure of train service $i$ from its preceding station before the connecting train service $j$ can depart. If $r_{ij}^{\text{max}}$ is the maximum (dwell or transfer) buffer time then the synchronization constraint is given as

$$\tau_j - \tau_i = t_r^i + t_{ij}^{\text{min}} + r_{ij} + z_{ij}T,$$

where $r_{ij} \in [0, r_{ij}^{\text{max}}]$ is the dwell/transfer buffer time, see Figure 1.

Crew and rolling stock connections give special synchronization constraints with respect to the logistics of crew and rolling stock schedules. First, a **turn** at a terminal station is a special stop where the minimum turn-around time does not just depend on alighting and boarding time, but also includes for instance the necessary time for cleaning the coaches and turning the locomotive. Second, a **crew connection** is a special transfer in which the crew of the feeder train is scheduled to transfer to the connecting train. In this case the connection can not be cancelled unless a reserve crew can be allocated to the connecting train. Third, a **rolling stock connection** is a connection in which trains are coupled/decoupled.

**Infrastructure constraints** describe headway restrictions due to the safety and signalling system and infrastructural limitations. Consider two trains services $i$ and $j$ that use a common track section some distance from the (distinct) departure stations. Let $t_r^i$ be the running time of train service $i$ to reach this point from its departure at the preceding station, and analog for $t_r^j$. Assume that a minimum headway $h_{ij}$ has to be respected from train service $i$ to $j$, and $h_{ji}$ if this sequence order is reversed. Then we have

$$h_{ij} \leq (\tau_j + t_r^j) - (\tau_i + t_r^i) - z_{ij}T \leq T - h_{ji}.$$

Rewriting in the form (2) gives the infrastructure constraint

$$h_{ij} + t_r^i - t_r^j \leq \tau_j - \tau_i - z_{ij}T \leq T - h_{ji} + t_r^i - t_r^j,$$

(3)
Figure 1: Synchronization constraint: the time window between the departure time $\tau_i$ of the feeder train and the departure time $\tau_j$ of the connecting train at the next station consists of the running time $t_i$, the minimum transfer time $t_{ij}^{\min}$, and the transfer buffer time $r_{ij}$.

So here $h_{ij} = h_{ij} + t_i - t_j$ and $r_{ij}^{\max} = T - h_{ij} - h_{ji}$, see Figure 2. The associated headway buffer time at the junction is thus $r_{ij} \in [0, T - h_{ij} - h_{ji}]$. Also the headway buffer time between train service $j$ and the next train service $i$ is here determined as

$$r_{ji} = r_{ij}^{\\max} - r_{ij},$$

since the sum of the minimum headways and headway buffer times of a train pair must equal the cycle time $T$.

Infrastructure constraints also arise with respect to simultaneous arrivals and departures at transfer stations, due to limitations on available train paths at the station track layout. Moreover, train services that use the same route to the next station are restricted by the safety distance of trains on these common track sections. This results in minimum arrival headways between arriving train pairs and minimum departure headways between departing trains. The associated constraints are also given as (3), where for the arrival headway constraints the running times $t_i$ and $t_j$ correspond to the running times from the preceding stations to the mutual station or to overtaking points at an intermediate stop, and for the departure headway constraints $t_i' = t_j' = 0$, see Figure 2.

Finally, additional market constraints are restrictions with respect to market requirements. This models for instance relative restrictions on the departure times of two train services to force semi-regular interdeparture times. For instance, if a line has twice the frequency of the overall cycle time then this line is modelled by two sequences of train services. The frequency constraint may then specify
that two train services have a semi-regular interdeparture time. Suppose that train services \( i \) and \( j \) must be separated by half the cycle time with a tolerance of \( \delta \) so that twice in a period a train service departs in a certain direction in a semi-regular fashion. Then the frequency constraint is

\[
\frac{1}{2} T - \delta \leq \tau_j - \tau_i - z_{ij} T \leq \frac{1}{2} T + \delta.
\]

Here \( l_{ij} = \frac{1}{2} T - \delta \) and \( r_{ij}^{\text{max}} = 2\delta \). The interpretation of the buffer time is in this case obsolete. It can now be interpreted as a tolerance where \( r_{ij}^{\text{max}} \) is the absolute tolerance. Market constraints may also specify exact departure times for a train service. This models prefixed departure times of for instance international (high speed) trains. For this a fictitious train service 0 is defined that has a predetermined departure time \( \tau_0 = 0 \). Assume that a train service \( j \) has to depart at exactly \( \tau_j = \tau \) then

\[
\tau \leq \tau_j - \tau_0 - z_{0j} T \leq \tau.
\]

The maximum buffer time is 0 min and so \( r_{0j} = 0 \) as well.

The time window constraint system can also be given in matrix notation. Let \( n \) be the total number of events, and \( m \) be the number of time window constraints. Number all constraints from 1 to \( m \). Suppose that the \( k \)th constraint is \( \tau_j - \tau_i = l_{ij} + r_{ij} + z_{ij} T \). Then define the matrix \( M \in \mathbb{Z}^{nxm} \) as \( (M)_{jk} = 1 \), \( (M)_{ik} = -1 \) and \( (M)_{lk} = 0 \) for all \( l \in \{1, \ldots, n\} \setminus \{i, j\} \). So a constraint corresponds with a column of two nonzeros. Let \( \tau \in [0, T]^n \) be the vector of all event times, and define vectors \( l \in \mathbb{R}^m \) and \( z \in \mathbb{Z}^m \) according to their constraint (arc) number, so
Figure 3: Constraint graph

e.g. \( l_k := l_{ij} \). Furthermore, denote the vector of buffer times as \( x \in \mathbb{R}_+^m \) defined as \( x_k := r_{ij} \), and equivalently \( \bar{x} \in [0,T]^m \) is defined as \( \bar{x}_k := \bar{r}_{ij}^{\text{max}} \). Then the constraint system is

\[
M^T \tau - Tz = l + x \\
x \in [0, \bar{x}], \quad z \in \mathbb{Z}^m.
\]

Note that we use the transpose matrix \( M^T \) rather than directly defining a matrix \( N = M^T \). This is motivated by the interpretation of the matrix \( M \) as the node-arc incidence matrix of a directed graph, the so-called constraint graph.

The constraint graph is a directed graph (or digraph) \( G = (V, E) \), where \( V \) is the set of \( n \) nodes and \( E \) is the set of \( m \) arcs, see Figure 3. The nodes \( i \in V \) correspond to the periodic events \( \tau_i \) and for each pair \( ij \) in a time window constraint there is a directed arc \( e_k = (i, j) \in E \) from \( i \) to \( j \). Each arc has an associated lower and upper capacity given as \( l_{ij} \) and \( u_{ij} = l_{ij} + r_{ij}^{\text{max}} \), respectively. Note that the constraint graph can have multiple arcs between two adjacent nodes, if several time window constraints are defined for one pair of event times. The columns of \( M \) define the arcs of the constraint graph: an arc \( e_k \) starts at node \( i \) for which \( (M)_{ik} = -1 \) and terminates at node \( j \) for which \( (M)_{jk} = 1 \).

In the sequel we assume that the constraint graph is strongly connected, i.e., there is a directed path between any two nodes. Otherwise, the problem can be decomposed into subproblems for which the constraint graphs are strongly connected. These subproblems can then be solved separately, and the solutions can afterwards be linked to one overall timetable. Note that the constraint graph contains all interactions between train services, including infrastructure restrictions. Second, we assume that the constraint graph does not contain loops, i.e., cycles of one arc. Note that a loop can not be defined by a node-arc incidence matrix. A loop would imply a relationship between an event to itself. This occurs for instance if a train circulates on a route without interactions to other trains but for one station. The loop constraint then specifies that the sum of the (circulation) running time and the dwell time at the one station where interactions occur has to be a multiple of the cycle time. Clearly, this constraint can be solved independently and does not influence the remaining constraint system.

The time window constraint system (6) contains both the event times \( \tau \) and the buffer times \( x \). Note that these variables are redundant: if we know the event times then we also know the resulting buffer times, and vice versa. Since we are
interested in optimizing the buffer times it is desirable to use these buffer times as decision variables. Note that since each buffer time $x_k \in [0, T]$, these variables do not suffer from modular arithmetic and can be used as variables in objective functions without causing discontinuities and thereby disturbing convexity. The next section derives an equivalent constraint system in the variables $x_k$ ($k = 1, \ldots, m$) only.

2.3 Cycle Constraints

Consider a simple railway network of two stations and trains circulating between the stations. Then the sum of running times and dwell times on this circulation must be a multiple of the cycle time since after returning at its origin station the train has to wait for its periodic departure time, which is exactly a multiple of an hour (or more generally, the cycle time) after its former departure, before starting a new circulation. So the sum of running times, minimum dwell times, and dwell buffer times has to be a multiple of the cycle time. Since the running times and minimum dwell times are given, this results in a constraint with respect to the buffer times. This is an example of a cycle constraint.

The above example can be generalized to circuits in a complex railway network with lots of interactions, represented by the constraint graph. From graph theory it is known that any circuit can be obtained by a linear combination of fundamental cycles [1]. A cycle is a sequence of adjacent arcs (a path) where the initial and final node is the same. The arcs on a cycle may have different directions. So it is not necessary that a cycle is an ordered sequence of successive processes (a circuit). Thus, more abstract, on each cycle in the constraint graph the oriented sum of the process times (running times, minimum dwell times, minimum transfer times, and minimum headway) and buffer times must be an integral multiple of the cycle time $T$. It is sufficient to consider only fundamental cycles of a cycle basis from which all other cycles (and all circuits) can be obtained. The (finite) number of fundamental cycles in the cycle basis, the so-called cyclomatic number, is $\nu = m - n + 1$, where $m$ and $n$ are the number of arcs and nodes in the graph, respectively. This can be clarified as follows: a spanning tree of a digraph $G$ is a subgraph that contains all nodes of $G$ but no cycles. The number of arcs of a spanning tree is $n - 1$ because any additional arc gives a cycle. The number of nontree arcs is therefore $m - (n - 1) = \nu$. Each nontree arc generates a fundamental cycle (together with the path on the spanning tree from the tail of the nontree arc to its head). The resulting $\nu$ cycles form a cycle basis. Note that a cycle basis is not unique.

The cycle constraint system defines all interactions between services given in terms of buffer times and taking into account the network structure. The periodicity of the event times is replaced by the periodicity of circulation times on cycles. Thus, the system of $\nu = m - n + 1$ cycle constraints is equivalent with the system of $m$ time window constraints (all interactions). Note that less cycle constraints are necessary to model all interactions as a result of incorporating the network structure. This implies that the number of integer variables in the
constraint system is reduced by \( n - 1 \).

A cycle constraint system can be obtained by using the cycle matrix representation of the constraint graph. The cycle (-arc) matrix \( \Gamma \) is defined as the \( n \times m \) matrix where each row is a vector representation of a fundamental cycle. That is, \( (\Gamma)_{ij} = 1 \) if arc \( j \) is a forward arc on cycle \( i \), \( (\Gamma)_{ij} = -1 \) if arc \( j \) is a backward arc on cycle \( i \), and \( (\Gamma)_{ij} = 0 \) otherwise. The direction of an arc on a cycle depends on the orientation of the cycle. Using the cycle matrix the constraint system becomes

\[
\begin{aligned}
\Gamma x - T z &= b \\
0 &\leq x \leq \bar{x}, \quad z \in \mathbb{Z}^n,
\end{aligned}
\tag{7}
\]

where \( \Gamma \in \mathbb{R}^{(n-1) \times m} \) is the cycle matrix, \( b \in \mathbb{R}^m \) is the vector of the negative oriented sum of minimum process times of the constraint graph/time window constraints, \( \bar{x} \in [0, T]^m \) is the vector of maximum buffer times, and \( T \in \mathbb{N} \) is the cycle time. Note that the right-hand side can be computed as \( b = -\Gamma l \), where \( l = (l_1, \ldots, l_m)^T \) are the minimum process times of the constraint graph/time window constraints. If a solution \( x \) to (7) has been computed then we can also simply compute the associated event time vector \( \tau \) by solving \( M^T \tau = l + x \) for \( \tau \), and subsequently set \( \tau \equiv \tau \pmod{T} \). Since \( l + x \) is now a given vector \( M^T \tau = l + x \) can be computed efficiently using standard numerical algorithms such as LU decomposition or Gauss-Jordan elimination. Moreover, note that we in fact only need \( n - 1 \) independent rows of \( M^T \) to completely determine \( \tau \) up to a constant (one degree of freedom). The event time vector \( \tau \) is then completely determined by fixing one event time. Note that if absolute constraints (5) are imposed then \( \tau \) is uniquely determined by fixing the fictitious event time \( \tau_0 = 0 \). On the other hand, if we have found a solution \( \tau \) to the time window constraint system (6) then the associated buffer time vector is simply obtained as \( x = M^T \tau - l \pmod{T} \). This shows that the time window constraint system (6) and the cycle constraint system (7) are equivalent.

Above we gave a graph-theoretical construction of a cycle basis based on a spanning tree in the graph. This gives an algorithm to compute the cycle matrix. Below we present an alternative algorithm for computing the cycle matrix based on linear algebra.

Recall from linear algebra that a null space of a matrix \( M \) is the space spanned by vectors \( x \) such that \( Mx = 0 \). A null space can be computed efficiently by Gaussian elimination with back-substitution (or by Gauss-Jordan elimination) in \( O(m^3) \) time for general matrices, see e.g. Strang [27]. The following theorem states that this algorithm can also be applied to compute a cycle matrix from the incidence matrix. Note that since the incidence matrix is very sparse (\( m = 2n \)), the efficiency of the algorithm is better than \( O(m^3) \).

**Theorem 1** Let \( G \) be a directed graph and \( M \) the associated (node-arc) incidence matrix. Then

1. a cycle basis of \( G \) is also a basis of the null space of \( M \),
2. if a basis of the null space of $M$ is given by vectors in $\{0, \pm 1\}^{m-n+1}$ then it is also a cycle basis.

3. the matrix of which the rows are the basis vectors of the null space of $M$ computed by Gaussian elimination with back-substitution, is a cycle matrix.

Proof. Consider the incidence matrix $M$ in more detail. The rows correspond to nodes and the columns to arcs in the associated digraph. At each row a 1 implies that the arc is an incoming arc of the node and a $-1$ implies an outgoing arc. On the other hand, each column in $M$ can be viewed as a vector representation of an arc, where the entry $-1$ denotes the tail and the entry 1 denotes the head. Now, observe that on a cycle the number of incoming arcs over all nodes equals the number of outgoing arcs. Thus, a cycle can be found as a linear combination of all columns (arcs) giving the zero vector, that is, by $x \in \{0, \pm 1\}^\nu$ solving $Mx = 0$. These vectors $x$ are just elements of the null space of $M$. On the other hand, the range of an incidence matrix $M$ is the space spanned by the vectors corresponding to the arcs in the graph, i.e., all possible combinations $Mx$ of the columns of $M$. A basis of the range is a set of independent vectors spanning this vector space. This corresponds to a spanning tree in the graph. The dimension of the range of a node-arc incidence matrix equals its rank, which is $n - 1$. The dimension of the null space then is $m - (n - 1) = m - n + 1$. Recall that a cycle basis also contains $\nu$ fundamental (independent) cycles. It follows that a cycle basis is also a basis of the null space of $M$. This proves the first statement of the theorem. Reversely, since a cycle is represented as a vector with entries in $\{0, \pm 1\}$ this is an additional requirement for a basis of the null space to be a cycle basis, which proves the second statement.

Finally, we prove the third statement. Assume that we compute $\nu$ linearly independent solutions $x$ to $Mx = 0$ by Gaussian elimination with back-substitution (choosing unit vectors for the free variables in the back-substitution process). Then $x \in \{0, \pm 1\}^m$. This can be proven as follows. The independent columns correspond to a spanning tree on $G$. Taking one free variable $x_k = 1$ at a time and the remaining free variables zero, and solving for the independent variables by back-substitution, gives the free variable (nontree arc) in terms of the independent variables which corresponds exactly to the associated path from the tail, across the spanning tree, to the head and closing the cycle. Or equivalently, the nontree arc combined with the reversed path on the spanning tree is zero. The tree arcs are traversed only once in the reversed direction and so $x = \{0, \pm 1\}^m$. □

2.4 Rolling Stock Constraints

The amount of buffer time to be scheduled in the timetable may be bounded by a maximum amount of rolling stock. This imposes additional constraints to the timetable constraint system. These constraints can be formulated as linear (in-)equalities of the buffer times and do not contain additional integer variables. An alternative way to include rolling stock constraints is to express them in terms of the integral multipliers $z$. In this way the rolling stock constraints restrict the
feasible region of the integral variables. This results in a smaller search space and can thus decrease the computational effort necessary to find the optimal solution.

Assume that trains circulate on fixed sequences of lines exclusively. An example is the popular case that trains are assigned to two lines corresponding to running a specific route in both directions. In this case a train arriving at its terminating station turns around to run in its opposite direction and is not assigned to a different line. Also more complex routings are possible covering a certain sequence of different lines that give a circuit in the service network. Such a circuit is called the **circulation** of a train.

A rolling stock constraint is modelled as follows. Let \( \bar{w}_i \) be a prescribed maximum number of trains on circulation \( i \). Then the rolling stock constraint is given as

\[
\sum_{k=1}^{m} u_{ik} x_k + d_i \leq \bar{w}_i T,
\]

where \( T \) is the cycle time, \( m \) is the number of connections/arcs in the constraint graph, \( x_k \) is the buffer time of stop \( k \) on the circulation, \( d_i \) is the minimum circulation time of circulation \( i \) (the sum of running times and minimum dwell and turn-around times), and \( u_{ik} = 1 \) if circulation \( i \) contains stop \( k \) and \( u_{ik} = 0 \) otherwise. Note that a train circulation is a circuit.

These constraints can also be written in vector notation. Let the number of circulations be denoted as \( v \) and define the matrix \( U \in \{0,1\}^{v \times m} \) as \( (U)_{ij} = u_{ij} \). Then the rolling stock constraints are given as

\[
U x + d \leq T \bar{w}. \tag{8}
\]

Here the inequality is defined componentwise. The vector \( d \) is given as \( d = Ul \), where \( l \) is the vector of minimum process times (or lower arc capacities in the constraint graph). The rolling stock constraints can now be taken into account by adding (8) to the periodic constraint system (7).

An alternative way to include the rolling stock constraints in the constraint system is to express the maximum number of trains of all circulations in the integral multipliers \( z \). This restricts the feasible region of the integral variables and the constraint system (7) is not expanded. The rolling stock constraints are now implicitly present in the feasible region of the integral variables.

Recall that any circuit can be written as a combination of fundamental cycles. The rows of the matrix \( U \) are the vectors representing the circulations, i.e., \( U = (u_1, \ldots, u_v)^T \), where \( u_i = (u_{i1}, \ldots, u_{im}) \). The cycle basis consists of the rows of the cycle matrix. A circulation \( i \) represented by the row vector \( u_i \) can thus be written as

\[
u_i = c_i \Gamma, \quad i = 1, \ldots, v, \tag{9}\]

where \( c_i = (c_{i1}, \ldots, c_{iv}) \) is a row vector representing the scalar multiples of the rows of the cycle matrix \( \Gamma \) (the fundamental cycles) which generate the circulation \( i \). Note that (9) can be written in matrix notation as \( U = C \Gamma \), where \( C = (c_1, \ldots, c_v)^T \in \mathbb{Z}^{v \times r} \) is an unknown matrix. This matrix \( C \) can be found by
solving (9) (or in standard form $u_i^T = \Gamma^T e_i^T$) for all $i = 1, \ldots, v$ using Gaussian elimination.

Now the feasible set of integral variables $z$ can be restricted as follows.

\[
\begin{align*}
Ux + d & \leq T\bar{w} \quad \iff \\
C\Gamma x + d & \leq T\bar{w} \quad \iff \\
C(b + Tz) + d & \leq T\bar{w} \quad \iff \text{(by (7))} \\
Cz & \leq \bar{w} - (Cb + d)/T.
\end{align*}
\]

This results in the following constraint on the feasible region:

\[Cz \leq \left[\bar{w} - \frac{1}{T}(Cb + d)\right].\]

In the same vein, a given minimum number $w_i$ of trains on all circulations $i$ results in the following constraint on the feasible region of the integral variables:

\[Cz \geq \left[w - \frac{1}{T}(Cb + d)\right].\]

Note that a minimum number of trains on a circulation can be obtained if all buffer times are zero. By (8) this gives $w \geq \lceil d/T \rceil$. So if no explicit minimum number of trains is given than this lower bound can be used to restrict the feasible integer region. Note that it is now easily checked whether or not a given minimum number of trains on a circulation is feasible by comparing it with this lower bound. If the number of trains $w$ on all circulations are fixed in advance then the feasible region is restricted by

\[Cz = w - (Cb + d)/T\]

Of course also combinations of the above three rolling stock circulation constraints are possible.

A different kind of a rolling stock constraint is obtained when the maximum amount of trains on the entire service network is given. In this case the allocation of trains to routes is still free. The optimization problem then also determines the optimal number of trains taking the given upper bound into account. Let $\bar{w} \in \mathbb{N}$ be the maximum number of trains. In the same notation as introduced above, the rolling stock constraint becomes

\[\sum_{i=1}^v \sum_{k=1}^m u_{ik} (x_k + l_k) \leq \bar{w} T.\]

This constraint implies that the sum of the minimum circulation times and buffer times on all circulations must not be larger than the maximum number of trains to be allocated multiplied by the cycle time. If a minimum amount of rolling stock is desired then the analog to (10) holds with $a \geq$ inequality, and for a fixed number of rolling stock the analog to (10) holds with equality.

Note that these rolling stock constraints may result in infeasibility if the number of trains is not enough to operate the service network according to a desired cycle
time $T$. This is easily checked by considering inequality (10) with all $x_k = 0$. The minimum necessary number of trains is thus at least

$$v \geq \left\lceil \sum_{i=1}^{v} \sum_{k=1}^{m} \frac{u_{ik}l_k}{T} \right\rceil.$$

Again the rolling stock constraint can also be expressed in the integral variables. In the same vein as before we then obtain for a given maximum amount of rolling stock

$$\sum_{i=1}^{v} \sum_{j=1}^{\nu} c_{ij}z_j \leq \left\lfloor \frac{\sum_{i=1}^{v} \sum_{j=1}^{\nu} (c_{ij}b_j) + \sum_{i=1}^{v} \sum_{k=1}^{m} (u_{ik}l_k)}{T} \right\rfloor.$$

This constraint gives an upper bound to a linear combination of the integral multipliers $z_j$ associated with the $j$th cycle constraints. For a given minimum amount of rolling stock the analog to (11) holds with a $\geq$ inequality (and rounding to the nearest integer above), and for a fixed amount of rolling stock the analog to (11) holds with equality. Note that the right-hand side of (11) is a fixed integer.
3 Railway Timetable Optimization

3.1 Introduction

The railway timetable optimization problem can be viewed as computing a feasible network timetable with an optimal allocation of buffer times. Buffer times are the variables in this optimization problem. The objectives differ between the various buffer times. Dwell buffer times should be as small as possible to obtain small travel times. Transfer buffer times can be used to guarantee reliable connections and so some buffer time has a positive impact. Turn-around buffer times should guarantee the stability of train circulations. Buffer times in rolling stock and crew connection times are convenient for improved punctuality. Headway buffer times reduce the probability of conflicts and are desired as large as possible.

This chapter presents the general periodic railway optimization problem based on the cycle constraints as derived in the previous chapter. These constraints agree with the DONS interface, that is, the generated files of variables and constraints used by CADANS (the solver of DONS) to compute a feasible timetable are also sufficient to obtain the variables and constraints in the optimization problem presented in this paper. So the optimization problem is compatible with DONS.

3.2 The Periodic Railway Timetable Optimization Problem

Let \( m \) be the total number of buffer times (or interactions), and let \( x_k \) \((k = 1, \ldots, m)\) be the buffer time with respect to an interaction \( k \). Then the periodic railway timetable optimization problem can be given as

\[
\min F(x_1, \ldots, x_m)
\sum_{k=1}^{m} \gamma_{ik} x_k - T z_i = b_i \quad i = 1, \ldots, \nu
\]

\[0 \leq x_k \leq \bar{x}_k \quad k = 1, \ldots, m
\]

\[z_i \in \mathbb{Z} \quad i = 1, \ldots, \nu.
\]

Here, \( F : \mathbb{R}^m \rightarrow \mathbb{R} \) is a cost function of the buffer times \( x_k \) \((k = 1, \ldots, m)\), \( \bar{x}_k \in [0, T] \) \((k = 1, \ldots, m)\) are the maximum buffer times, \( T \) is the cycle time (or period length), \( -b_i \) is the oriented sum of minimum process times on a fundamental cycle \( i \) of the constraint graph, \( \nu \) is the number of fundamental cycles in the constraint graph, \( z_i \) \((i = 1, \ldots, \nu)\) are integers, and \( \gamma_{ik} = 1 \) if interaction \( k \) corresponds to a forward arc on cycle \( i \), \( \gamma_{ik} = -1 \) if interaction \( k \) corresponds with a backward arc on cycle \( i \), and \( \gamma_{ik} = 0 \) otherwise. Note that the (cycle) constraints state that the oriented sum of process times and buffer times on a fundamental cycle of the constraint graph equals an integral multiple of the cycle time \( T \).
The optimization problem can also be formulated in vector notation as
\[
\min F(x) \\
\Gamma x - Tz = b \\
0 \leq x \leq \bar{x}, \quad z \in \mathbb{Z}^r,
\]
(13)
where \(\Gamma \in \mathbb{R}^{n \times m}\) is the cycle matrix, \(b \in \mathbb{R}^m\) is given as \(b = -\Gamma l\) where \(l \in [0, \infty)^m\) is the vector of the minimum process times of the constraint graph/time window constraints, \(\bar{x} \in [0, T]^m\) is the vector of maximum buffer times, and \(T \in \mathbb{N}\) is the cycle time.

Additionally, the integers can be constrained by a convex set \(\Omega \subseteq \mathbb{Z}^r\) defined by constraints on the amount of rolling stock, see Section 2.4. Also integer lower and upper bounds can be computed, see Section 3.4. These integer bounds can also be incorporated in \(\Omega\).

The function \(F : \mathbb{R}^m \to \mathbb{R}\) models the ‘cost’ of any allocation of buffer times over the timetable. From a computational point of view the cost function \(F\) has to be convex implying that any local optimal solution is also the global optimum. For nonconvex \(F\) the problem is practically unsolvable. A convenient cost function is a (weighted) sum of the relative costs of all buffer times, i.e.,
\[
F(x) = \sum_{k=1}^{m} c_k F_k(x_k)
\]
where \(c_k \geq 0\) are positive real numbers, and the functions \(F_k : \mathbb{R} \to \mathbb{R}\) are smooth convex functions of a single buffer time. The weights \(c_k\) give the relative importance of the individual cost contributions and thereby tune the priorities of the buffer times. This particular form of \(F\) is a convex separable function of the buffer times. Note that a linear cost function \(c^T x\) is a special case of (14). This separable formulation is convenient for modelling the individual cost contributions of all buffer times. In the next chapter several appropriate cost functions are presented representing the buffer time costs to passengers and/or railway operators.

Another possible formulation of cost functions is \(F(x) = \max_{k=1, \ldots, m}[F_k(x_k)]\), resulting in a minimax problem. The objective is here to minimize the maximum of all buffer time costs. For example, if \(F_k(x_k) = x_k\) \((k = 1, \ldots, m)\) then the objective is to minimize the maximal buffer time over all buffer times. However, for our purposes this objective is not felt appropriate as the individual buffer times vary in interpretation. Only for special cases of uniform variables this objective may be interesting. For instance, if we are only interested in minimizing the transfer times over the network. Also the priorities of the individual buffer times can not be modelled in a minimax problem.

3.3 Computational Complexity

The minimization problem (12) is a convex mixed-integer programming problem, with linear mixed-integer constraints and a convex (separable) cost function. Note
that although the arrival/departure times are periodic events, the buffer times $x_k \in [0, T]$ are not periodic. Therefore the cost function is not affected by periodicity. If the optimization problem would have been formulated directly with respect to the time window constraints and the cost function as a function of the periodic arrival/departure times then the cost function would have been non-convex by the modular arithmetic (arrival and departure times have to be evaluated modulo the cycle time).

For fixed integers $z$ (and $F$ convex in $x$), the optimization problem (12) becomes a convex programming problem, and if $F$ is linear in $x$ then the problem is a linear programming problem. These problems can be solved efficiently.

However, the integer variables $z$ in the mixed-integer constraints make the problem difficult to solve. In fact, the optimization problem (12) is $NP$-complete which follows from the $NP$-completeness of the underlying feasibility problem of finding solutions to the cycle constraint system (or the time window constraint system). For small networks the railway optimization problem can be solved exactly by a branch-and-bound algorithm [21] using convex programming relaxations.

For large-scale networks more effort is necessary to solve the problem. Both for solving the convex relaxation problem and the mixed-integer part. In particular the structure of the timetable constraint system (the timetable polyhedra) can be examined for preprocessing, and to tighten the feasible region by which the branch-and-bound procedure becomes more effective. Nachtigall [20] examined the timetable polyhedron and found two classes of facet defining valid inequalities. Moreover he developed two polynomial separation algorithms which can be used to generate strong cutting planes.

### 3.4 Preprocessing: Integer Bounds

Solving the constraint cycle system is complicated by the integrality condition on the vector $x$. However, lower and upper bounds on the integral variables $z_i$ can be obtained by exploiting the network structure. This reduces the search space considerably.

Recall that $x \in [0, \bar{x}]$. Then a lower bound on a cycle is obtained if all backward arcs have a maximum buffer time and the forward arcs have zero buffer time. Likewise, an upper bound is given if all forward arcs have a maximum buffer time and the backward arcs have zero buffer time. Consider a cycle $i$ in the constraint graph. Define $\gamma_{ik}^+ = \max(0, \gamma_{ik})$, i.e., $\gamma_{ik}^+ = 1$ if arc $k$ is a forward arc on cycle $i$, and $\gamma_{ik}^+ = 0$ otherwise. Equivalently, define $\gamma_{ik}^- = \min(0, \gamma_{ik})$, i.e., $\gamma_{ik}^- = -1$ if arc $k$ is a backward arc on cycle $i$, and $\gamma_{ik}^- = 0$ otherwise. Then

$$\sum_{j \neq k} m \gamma_{ik}^- x_k \leq b_i + x_i T \leq \sum_{k=1}^m \gamma_{ik}^+ x_k.$$

Let the $\nu \times m$ matrices $\Gamma^+$ and $\Gamma^-$ be defined accordingly as $(\Gamma)^{+}_{ij} = \gamma_{ij}^+$, and
Then we obtain the lower bound vector
\[
\overline{z} := \left\lfloor \frac{1}{T} (\Gamma^\omega \overline{x} - b) \right\rfloor
\]
and upper bound vector
\[
\overline{z} := \left\lceil \frac{1}{T} (\Gamma^+ \overline{x} - b) \right\rceil.
\]
Here \([a]\) is the nearest integer above \(a\) (the ceiling operator) and \([a]\) is the nearest integer below \(a\) (the floor operator), defined componentwise. An additional constraint to the cycle constraint system \((7)\) is thus
\[
z \leq z \leq \overline{z}.
\]
Note that the choice of the cycle basis influences the lower and upper bounds on the integral vector \(z\), and thus also the solution algorithm efficiency.

The above derived integer bounds depend on the decision variable upper bounds \(\overline{x}\). These upper bounds (maximum buffer times) may be taken equal to the cycle time \(T\). In this way the overall optimal solution results. Note that the upper bounds have become somewhat superfluous as large buffer times are penalized by the cost functions. However, removing the buffer time upper bounds magnifies the gaps between the lower and upper bounds on \(z\) given by \((15)\) and \((16)\). In a branch-and-bound scheme this implies larger computation times. On the other hand, tight upper bounds \(\overline{x}\) on buffer times to avoid large waiting times may result in infeasibility of the constraint system.
4 The Network Synchronization Problem

4.1 The Optimization Problem

The aim of this paper is to find an optimal network synchronization. So we are concerned with buffer times between train connections. These include buffer times at stops, transfers, rolling stock connections, crew connections, and turns. The resulting optimization problem will be referred to as the network synchronization problem.

We consider the following optimization model as obtained in the two previous chapters:

\[
\min \sum_{k=1}^{m} c_k F_k(x_k)
\]

\[
\sum_{k=1}^{m} \gamma_{ik} x_k - T z_i = b_i \quad i = 1, \ldots, \nu
\]

\[
0 \leq x_k \leq \bar{x}_k \quad k = 1, \ldots, m
\]

\[
z_i \in \mathbb{Z} \quad i = 1, \ldots, \nu.
\]

where \(c_k \geq 0\) are positive real numbers, and the functions \(F_k : \mathbb{R}_+ \rightarrow \mathbb{R}\) are smooth convex functions of a single buffer time \(x_k \geq 0\). Additionally, the integer vector \(z\) may be constrained by a feasible region \(\Omega\) defined by the rolling stock constraints (Section 2.4) and the integer bounds (Section 3.4). In this chapter we will present some appropriate cost functions \(F_k\) for the network synchronization problem.

We are concerned with an optimal allocation of buffer times over the network. Therefore, it is convenient to express cost in terms of time (minutes). Of course other cost units are possible as for instance the cost equivalent in cash. The following cost functions are proposed:

- **Buffer time**: minimizing dwell buffer times to obtain small travel times.
- **Expected departure delay**: maximizing buffer time reliability of logistics connections (crew connections, rolling stock connections, and turns of trains at terminal stations (stability of train circulation)).
- **Expected transfer waiting time**: optimizing a trade-off between a small transfer buffer time and a high transfer reliability.

The cost function \(F\) in the optimization problem is a (weighted) sum of these cost functions. The weights give the relative importance of the individual cost contributions. This tunes the priorities of circulation stability, crew and rolling stock connections, transfers, and travel times. The resulting network synchronization problem aims at finding a timetable with small travel times and transfer waiting times, and a punctual and reliable railway operation.
Also other choices of cost functions are possible representing the specific objectives of the problem at hand. Recall that we are concerned with connections only and so we did not consider cost functions for headway buffer times. Of course, appropriate cost functions can be derived that penalize small headway buffer times between a pair of trains in both directions. The subsequent sections consider the individual cost functions for the network synchronization problem.

### 4.2 Tight Connections

An obvious measure for the quality of a timetable is the total (synchronization) buffer time, or the total passenger waiting time resulting from these buffer times. Optimizing this measure results in a linear cost function to the network synchronization problem: the weighted sum of buffer times. So the linear cost function is

\[ F(x) = \sum_{k=1}^{m} c_k x_k, \]

where \( c_k \geq 0 \) is a constant weight reflecting the priority of dwell/transfer connection \( k \), for instance an estimate of the (relative) passenger flow, and \( c_k = 0 \) for headway buffer times.

The solution to the linear network synchronization problem gives an optimal timetable for punctual operation. However, with respect to arrival delays the resulting timetable may be far from optimal. The cost function (19) tends to give tight transfer times for high priority connections. In case of arrival delays this implies that either the connecting train is frequently delayed as well or, in the case of transfer connections, large amount of transferring passengers miss their connection. The linear cost function for all buffer times should therefore be considered only if the arrival times have a large punctuality percentage.

However, for dwell buffer times \( x_k \) the linear cost function \( F_k(x_k) = c_k x_k \) is convenient as it minimizes passenger travel times and train circulation times. Note that train running times include running time margins to neutralize the stochastic driving behaviour. Dwell buffer times also compensate arrival delays but this is just a (positive) side-effect. In principle the dwell times should be as small as possible and possible variation in running times should be compensated by running time margins.

An interesting observation is that the linear network synchronization problem (where \( F(x) = c^T x \)) has integral optimal buffer times as a result of a special property of the cycle matrix and the node-arc incidence matrix (total unimodularity).

**Theorem 2** Consider the (linear) network synchronization problem (18) with linear cost function (19). If all parameters \( T, b_k, \bar{e}_k \) \( (k = 1, \ldots, m) \) are integral then the linear network synchronization problem has integral solutions, i.e., the optimal buffer times and corresponding event times are also integers.
Proof. Note that $b = \Gamma l$ is integral if $l$ is, since then $b$ is just a linear combination of integral vectors. Moreover, since $z \in \mathbb{Z}^r$ and $T \in \mathbb{Z}$ also $zT \in \mathbb{Z}^r$. From integer programming theory it is known that for a totally unimodular (TU) matrix $A$ the solution (if any) of the linear programming problem $\{\min x | Ax \leq b, x \in \mathbb{R}_+^m\}$ is integral [21]. So in particular the relaxation problem of an MIP has integral solutions, and thus also the MIP itself. Now, the cycle time is TU which can be proven from its construction of which all operations are closed under total unimodularity. Note that the proof also holds for the network synchronization problem with time window constraints. The matrix in this case is just the transpose of the incidence matrix which is totally unimodular.

Note that it is convenient that all arrival and departure times in a timetable are given in minutes (integers). Theorem 2 then states that this is automatically obtained by the optimal solution of the linear network synchronization problem.

4.3 Reliable Connections

A major cause of (secondary) delays is represented by crew connections in which a driver or conductors have to change trains, and by rolling-stock connections for, e.g., coupling of coaches. These connections are hard, that is, a connecting service can only depart after the processes of the connection have been completed. For these connections the reliability of the buffer time between the arriving and departing services is hence very important. A third category to which buffer time reliability is of major concern is the stability of train circulations. The desire of a punctual initial departure time of a new train circulation has to be assured by a buffer time in the turn-around time at the terminating station. Note that running time margins and dwell buffer times during the circulation also compensate for possible delays at the cost of increased (passenger) travel times. A relative large buffer in the turn-around time does not influence travel times but assures a punctual departure at the start of a new circulation. In general, a reliable connection has to be assured for logistics constraints.

A measure for the reliability of a buffer time is the expected departure delay of the connecting service due to failure of the buffer time. The expected departure delay is formally given as follows. Consider a connection $k = (i, j)$ at a railway station from a train service $i$ to a subsequent train service $j$. Let $t_i^*$ be the scheduled running time of service $i$ from the preceding station, and $x_k \geq 0$ be the connection buffer time. Furthermore, let $T_i$ be a random variable denoting the stochastic running time of service $i$, with density function $g_i$ and associated distribution function $G_i(t) = \int_0^t g_i(\tau)d\tau$, see Figure 4. Common distributions are the shifted Weibull, lognormal, Erlang, or Gamma distributions, but also empirical density functions may be used. Denote the stochastic departure delay of train service $j$ as $Y_j$. Then the probability that train $j$ is delayed is $P(Y_j > 0) = P(T_i > t_i^* + x_k)$. Thus, the expected departure delay is $E[Y_j] = E[(T_i - t_i^* - x_k)^+]$ with $(\cdot)^+ = \max(0, \cdot)$, that is

$$E[Y_j] = \int_{t_i^*+x_k}^{\infty} (\tau - t_i^* - x_k)g_i(\tau)d\tau$$
Note that the integration variable \( p \) corresponds to the arrival delay of service \( i \). So instead of the running time distribution, also the distribution of the induced stochastic arrival delay \( \text{P} i \) of service \( i \) is sufficient. Note that the probability that train service \( i \) arrives on-time is given as \( \text{P}(T_i \leq t'_i) = G_i(t'_i) \). So typically, a random arrival delay \( \text{P}_i \) has a density function

\[
 f_i(p) = \begin{cases} 
 G_i(t'_i) & \text{if } p = 0 \\
 g_i(p + t'_i) & \text{if } p > 0,
\end{cases}
\]

where \( g_i, G_i, \) and \( t'_i \) are interpreted as above.

Above we silently assumed that the arrival delay of a feeder train is the result of a stochastic running time only. Thus, a possible departure delay at the start of the train run is neglected. The underlying idea is that at each connection the buffer time is computed to cope with the primary delays as good as possible. But this also implies that the secondary delays are controlled as good as possible. Therefore, this approach can be interpreted as the minimization of a first-order approximation of the delay propagation with respect to primary delays.

Instead of looking at the expected departure delay we can more generally look at the expectation that an event is \( \delta \) minutes later, with \( \delta \geq 0 \). The probability that train \( j \) departs \( \delta \) minutes late is given as \( \text{P}(Y_j \geq \delta) = \text{P}(T_i \geq t'_i + x_k + \delta) \) and the expectation of a departure delay more than \( \delta \) minutes is \( \text{E}[(T_i - t'_i - x_k - \delta)] \), i.e.,

\[
 \text{E}[Y_j - \delta] = \int_{t'_i + x_k + \delta}^{\infty} (\tau - t'_i - x_k - \delta)g_i(\tau)d\tau \\
 = \int_0^{\infty} (p - x_k - \delta)^+g_i(p + t'_i)dp.
\]

The expected departure delay is now a special case for \( \delta = 0 \). The following theorem states that the expectation \( \text{E}[Y_j - \delta] = \text{E}[P_i - x_k - \delta] \) is convex in the buffer time \( x_k \) for all \( \delta \geq 0 \).
Theorem 3 Let \( g : [0, \infty) \to [0, \infty) \) be a density function, and \( t, \delta \in [0, \infty) \). Then
\[
\int_0^\infty (p - x - \delta)^+ g(p + t) \, dp
\]
is a convex function of \( x \) on the domain \([0, \infty)\).

Proof. Recall that \( \max(0, x) \) is a convex function. Denote the integral function (20) as \( h(x) \). Let \( \alpha \in [0,1] \). Then we have to prove that \( h(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha h(x_1) + (1 - \alpha)h(x_2) \). We have
\[
h(\alpha x_1 + (1 - \alpha)x_2) = \int_0^\infty (p - \alpha x_1 - (1 - \alpha)x_2 - \delta)^+ g(p + t) \, dp
\]
\[
= \int_0^\infty (\alpha(p - x_1 - \delta) + (1 - \alpha)(p - x_2 - \delta))^+ g(p + t) \, dp
\]
\[
\leq \int_0^\infty [\alpha(p - x_1 - \delta)^+ + (1 - \alpha)(p - x_2 - \delta)^+] g(p + t) \, dp
\]
\[
= \alpha \int_0^\infty (p - x_1 - \delta)^+ g(p + t) \, dp + (1 - \alpha) \int_0^\infty (p - x_2 - \delta)^+ g(p + t) \, dp
\]
\[
= \alpha h(x_1) + (1 - \alpha)h(x_2).
\]

This cost function can also be applied to optimize running time margins. So let the running time also consist of a constant minimum running time and a variable running time margin that has to be determined in the optimization. Then the running time margins are also decision variables in the optimization problem. Now the running time margin and the dwell buffer time (or turn-around buffer time, etc.) are jointly optimized. An appropriate cost function for the running time margin is then for instance the expectation of arriving 2 min late, and for the dwell buffer time just the buffer time. Then the optimization problem aims at optimizing the trade-off between a reliable running time (margin) and a small dwell (buffer) time. Note that more flexibility of the process times represented by the variable (buffer time/margin) part, also reduces the chance of infeasible constraints.

A further extension of the optimization model is to include knock-on delays along the route of a train. Carey and Kwieciński proved that the resulting expected delays along the route are convex in the successive buffer times/margins. However, the evaluation of these cost functions may take a lot of computation time. So here is a trade-off between a more accurate model and tractability of the optimization problem. A better approach may be to use these stochastic models in a post-evaluation of the solution to the optimization model.

4.4 Transfer Connections

The effect of arrival delays on transfer reliability can be dealt with by incorporating the risk and significance of missing connections in the objective function. This can be accomplished by considering the mean or expected transfer waiting
Consider a connection $k = (i, j)$ from a train service $i$ to $j$. The transfer waiting time is a function of the arrival delay $p_i$ of the feeder service $i$. If the feeder service arrives on time or within the transfer buffer time $x_k$ then the transferring passengers have to wait for the scheduled departure time. For intermediate arrival delays of the feeder service the connecting service $j$ may wait to secure the connection, and departs right after the arrival of the transferring passengers. If the feeder service has an arrival delay for which securing the connection would result in exceeding the maximum admissible train waiting time, the synchronization control margin $\bar{s}_j$, then the connection is cancelled and the transferring passengers miss the connection. The passengers then have to wait on the next train of the connecting line. Assuming that the connecting train departs on-time, the transfer waiting time can be given as

$$w_k(p_i) = \begin{cases} 
  x_k - p_i & \text{if } 0 \leq p_i \leq x_k \\
  x_k + \omega_j - p_i & \text{if } x_k + \bar{s}_j < p_i \leq x_k + \omega_j \\
  0 & \text{otherwise},
\end{cases}$$

(21)

see Figure 5. Here $x_k$ is the transfer buffer time of connection $k = (i, j)$, $\omega_j$ is the interdeparture time (the reciprocal of the frequency) of the connecting service $j$, and $\bar{s}_j < \omega_j$ is the synchronization control margin of train $j$.

The above derivation of the transfer waiting time assumes that transferring passengers who miss the connection are able to catch the next train of the connecting line and also do not leave the station in a train of an alternative line. Moreover it is assumed that the connecting train and its successor depart on time. These assumptions are reasonable from a scheduling point of view and result in a considerable reduction of complexity.

To obtain an expression for the expected transfer waiting time the distribution

![Figure 5: Transfer waiting time](image)
of the arrival delays is required. In the sequel it is assumed that early arrivals do not contribute to transfer waiting time costs, i.e., early arrivals are treated as on-time arrivals.

Let \( g_i \) be the density function of the stochastic running time \( T_i \) of the feeder train service \( i \) and \( G_i(t) = \int_0^t g_i(\tau) d\tau \) the corresponding distribution function. Let \( t_i^s \) be the scheduled running time. The probability that the train arrives on-time is simply \( P(T_i \leq t_i^s) = G_i(t_i^s) \). The delay density function is therefore

\[
f_i(p) = \begin{cases} 
G_i(t_i^s) & \text{if } p = 0 \\
g_i(p + t_i^s) & \text{if } p > 0.
\end{cases}
\]

The following theorem gives a sufficient condition for which the resulting expected transfer waiting time is a convex function.

**Theorem 4** Consider a transfer connection \( k = (i,j) \). Let \( w_i > \bar{s}_j \geq 0 \), the function \( w_k(p) \) defined as in (21), and \( P \geq 0 \) be a stochastic variable with a nonincreasing differentiable density function \( f_i \) with a possible discontinuity at \( p = 0 \). Then the expectation

\[
E[w_k(P)] = x_k P(p = 0) + \int_{0^+}^{\infty} w_k(p) f_i(p) dp
\]

is a convex function of \( x_k \geq 0 \).

**Proof.** Recall that \( 0^+ = \lim_{\epsilon \downarrow 0} \epsilon \) is necessary because of a possible discontinuity of \( f_i \) at 0. Let \( \bar{w}_k(x_k) = E[w_k(P)] \). Below it is shown by construction that the second derivative of \( \bar{w}_k \) exists and is nonnegative which proves convexity. Let \( F_i \) be the distribution function associated with \( f_i \). Note that \( f_i \) is integrable as it is a density function and \( \bar{S}_i \) exists. Moreover note that \( F_i(0) = P(p = 0) \). Using partial integration the expectation of \( w_k(P) \) can be rewritten as

\[
\bar{w}_k(x_k) = x_k P(p = 0) + \int_{0^+}^{x_k} (x_k - p) f_i(p) dp + \int_{x_k}^{x_k + h_j} (x_k + h_j - p) f_i(p) dp
\]

\[
= x_k P(p = 0) + \lim_{\epsilon \downarrow 0} \left( [(x_k - p) F_i(p)]_{\epsilon}^{x_k} + \int_{\epsilon}^{x_k} F_i(p) dp \right)
\]

\[
+ [(x_k + h_j - p) F_i(p)]_{x_k + \bar{s}_j}^{x_k + h_j} + \int_{x_k + \bar{s}_j}^{x_k + h_j} F_i(p) dp
\]

\[
= x_k P(p = 0) - x_k F_i(0) + \int_{0^+}^{x_k} F_i(p) dp
\]

\[
- (h_j - \bar{s}_j) F_i(x_k + \bar{s}_j) + \int_{x_k + \bar{s}_j}^{x_k + h_j} F_i(p) dp.
\]

The first derivative of \( \bar{w}_k \) can be computed using the fundamental theorem of calculus, i.e., \( \frac{d}{dx} \int_{a}^{x} F(p) dp = F(x) \) for any fixed \( a \leq x \). Hence,

\[
\bar{w}_k'(x_k) = F_i(x_k) - (h_j - \bar{s}_j) f_i(x_k + \bar{s}_j) + F_i(x_k + h_j) - F_i(x_k + \bar{s}_j).
\]

---

Improving Punctuality and Transfer Reliability by Railway Timetable Optimization
By the assumption of $f_i$ being nonincreasing, we obtain for the second derivative of $\bar{W}_k$

$$\bar{W}''_k(x_k) = \begin{cases} f_i(x_k) - f_i(x_k + \bar{s}_j) - (h_j - \bar{s}_j) f'_i(x_k + \bar{s}_j) + f'_i(x_k + h_j) & \geq 0 \\ \geq 0. & >0 \end{cases}$$

Typically, railway planners (unconsciously) schedule the running time after the top of its density function by which most trains (e.g. 87%) arrive on-time, see Figure 4. The expected transfer waiting time is then a convex function by Theorem 4. Examples of convenient distributions are the Weibull and Erlang distribution, or any empirical density function with a decreasing tail, see Figure 4. The expected transfer waiting time can be computed numerically. Goverde [9] considers a negative-exponential density for the arrival delays with an additional punctuality rate. This can be viewed as approximating the tail of the running time density by a negative exponential density. For this case an analytical expression for the expected transfer waiting time can be computed [9], which is convenient in the overall optimization problem.

Figure 6 shows a typical expected transfer waiting time function of the transfer buffer time for some values of the synchronization control margin. Note that the function is convex. For large buffer times (larger than 8 minutes) the influence of the synchronization control margin is negligible: any arrival delay can be compensated by the buffer time. On the other hand, the effect of the synchronization control margin is considerable for small buffer times. If the synchronization control margin is large (about 5 minutes and larger) then the minimum of the expected transfer waiting time is obtained for zero transfer buffer time. The synchronization control can then compensate all (mainly occurring) arrival delays.
Conclusions

The computation of a periodic railway timetable can be modelled conveniently as an optimization problem, where the buffer times are the decision variables and the cost function is convex, by taking the network structure of the (periodic event time window) constraints in account. The resulting MIP constraint system of the buffer times is compatible with the DONS system of the NS, whereby the optimization problem is suitable to extend the DONS features.

An appropriate convex cost function of the buffer times can be derived representing the individual required performance of a tight connection, a reliable connection, or a transfer connection, and evaluating the priorities between the individual buffer time performance costs. The convex cost function guarantees that a local optimal solution gives the global optimum as well.

Using the network structure of the constraints in the problem formulation significantly reduces the required number of integer variables as opposed to the conventional periodic event (time window) constraint system. The network structure is also used in CADANS to solve the time window constraint system [25]. However, modelling the network structure explicitly gives opportunities to restrict the feasible integer region by integer bounds, rolling stock constraints, and strong cutting planes. This is advantageous for solving large-scale problems.

Another progress is made by using cost functions to penalize large buffer times at stops, connections, and turns, making upper bounds on these buffer times superfluous. This considerably increases the opportunity to find feasible solutions, and thus implicitly solves the problem of finding conflicting constraints. Moreover, also running time margins can be included as decision variables in the optimization problem which improves the feasibility issue as well as timetable efficiency even more.

Current research concentrates on efficient solution algorithms to solve the presented optimization problems. Moreover, the optimization approach will be validated by a case study. Experiments with small networks give promising results.

The proposed optimization model contributes to design efficient and robust railway timetables to improve performance and attractiveness of railway operations, increase service quality to passengers, and decrease costs to railway operators and passengers due to delays.

Acknowledgements

This publication is a result of the research programme Seamless Multimodal Mobility (SMM), carried out within the TRAIL Research School for Transport, Infrastructure and Logistics, and financed by Delft University of Technology.
References


Travel Choice Behavior in Public Transport Networks
A Fuzzy Genetic Approach

TRAIL Research School, Delft, December 1999

Authors
Ir. Sascha Hoogendoorn-Lanser and Dr.ir. Serge P. Hoogendoorn
Faculty of Civil Engineering and Geosciences, Delft University of Technology

Thesis supervisor
Prof.dr.ir. Piet H.L. Bovy
Faculty of Civil Engineering and Geosciences, Delft University of Technology


## CONTENTS

### ABSTRACT

1 INTRODUCTION
  1.1 Research motivation ........................................... 1
  1.2 Research objective ............................................. 1
  1.3 Research approach ........................................... 1
  1.4 Theoretical and practical results ........................... 1
  1.5 Outline of paper .............................................. 2

2 MODELING MULTI-MODAL TRAVEL CHOICE BEHAVIOR .......................... 3
  2.1 Traditional modeling approaches ............................. 3
  2.2 Motivation for using fuzzy logic and fuzzy sets ............ 3

3 OUTLINE OF FLIRT-MODEL .............................................. 6

4 FUZZY SET DESCRIPTION OF ATTRIBUTES AND UTILITIES ...................... 8
  4.1 Modeling trip attributes using fuzzy sets .................. 8
  4.2 Crisp sets versus fuzzy sets .................................. 8
  4.3 Fuzzy trip chain attributes .................................. 9
  4.4 Fuzzification of trip chain attributes ...................... 10
  4.5 Fuzzy utilities ............................................... 10

5 APPROXIMATE REASONING USING FLIRT-MODEL .................................. 11
  5.1 Fuzzy decision rules .......................................... 11
  5.2 Fuzzy inference and approximate reasoning .................. 12
  5.3 Rule base choice .............................................. 12

6 CHOICE OF TRIP ALTERNATIVE ........................................... 14
  6.1 Defuzzification and comparison of crisp utilities ........... 14
  6.2 Fuzzy ranking .................................................. 14
  6.3 Direct comparison of fuzzy utilities .......................... 15

7 CALIBRATION OF FLIRT-MODEL ........................................... 16
  7.1 Performance measure .......................................... 16
  7.2 Choice of membership functions .............................. 17
  7.3 Calibration of FLiRT-model using a Genetic Algorithm ...... 17
Abstract

Travel choice behavior models are valuable tools providing insight into travel choice behavior that is required in among others the design and planning of transport systems. In this paper, we propose a new fuzzy approach to describe travel choice behavior in urban public transport networks.

We motivate the fuzzy logic approach, since it enables handling traveler's erroneous and inexact perception and vague appraisal of trip chain attributes. Moreover, traveler's knowledge and experience are reflected in an interpretable rule base consisting of a limited set of fuzzy decision rules. Travel choice decisions are made based on these rules. Furthermore, the model provides mechanisms to include non-linear, partially compensational, and partially lexicographic characteristics of the travel choice process.

Besides providing a more realistic representation of human decision processes, the proposed fuzzy approach is mathematically tractable. Among other things, this implies that calibration, which is a common problem of fuzzy approaches, is possible.

The fuzzy model resulting from calibration with a genetic algorithm is both easy to interpret as well as insightful. Comparative analysis of the developed model and binomial logit models reveals that the predictive performance of the developed model significantly improves compared to traditional (binomial logit) models.
1 Introduction

In this introduction, we summarize the motivation, the objectives, and the approach of the research presented in this paper. Moreover, we outline both the theoretical and practical results presented in the ensuing.

1.1 Research motivation

Insight into choice behavior and travelers’ appraisal of trip chain attributes is of dominant importance in both the planning of multi-modal transport systems, as well as market assessment for new public transport services (e.g. light-rail). To obtain these insights, travel choice models can be used.

However, travel choice models based on random utility theory are not able to predict travel choice sufficiently accurate. We argue that among the causes for this poor performance are imprecisions and the qualitative nature of travelers’ appraisal of observable and unobservable trip attributes. Additionally, random utility models assume explicit relations for both the systematic utility and the distribution function of the random utility component present. The systematic utility is determined from the attributes of the trip chain based on the concept of trade-off. However, it is not likely that such explicit relations realistically describe human decision mechanisms, nor that compensational decision mechanisms are applicable in each choice situation. In fact, human decision processes are known to be highly non-linear, partially compensational, and partially lexicographic. Finally, travel choice behavior is not simply a result of rational utility maximization. Travelers tend to make travel choices based upon past experiences, knowledge, and habit as well.

1.2 Research objective

The objective of the research is to develop a route choice model for public transport networks with sufficient predictive capabilities to be useful in planning and market assessment. Moreover, human appraisal and decision making mechanism are to be modeled realistically.

1.3 Research approach

To improve model performance, we propose a new model based on the fuzzy logic paradigm. Since fuzzy sets are very well suited to describe imprecise and linguistic trip chain attributes, the developed modeling approach can adequately deal with travelers’ imprecise appraisal of trip chain attributes. Moreover, the relation between the trip chain attributes and the utility is not a-priori constrained. Moreover, we argue that experiences and knowledge of travelers can be captured in a limited set of decision rules, the so-called rule base. Trip chain alternatives are compared to experienced situations, after which a decision is made using approximate reasoning. In our approach, only travel-time components of urban public transport are considered.

1.4 Theoretical and practical results

We can conclude that the developed fuzzy model does not only theoretically improve upon traditional models considering the human perception and deci-
sion process, but also yields significant improvements in predictive performance. As we will show using data collected in Dutch public transport networks, the developed model correctly predicts the preferred alternative in 84% of all cases. In comparison, traditional binomial logit models only predict the preferred trip alternative in 72% of all cases.

1.5 Outline of paper

This paper is outlined as follows. In the second section, we will discuss traditional travel choice models for multi-modal travel choice processes. Additionally, we will motivate in detail why a fuzzy approach was chosen. In section 3, we will present an overview of the fuzzy model. Section 4 describes modeling of trip attributes and utilities using fuzzy sets. Section 5 discusses approximate reasoning using fuzzy logic to determine fuzzy utilities of the trip alternatives. In section 6 we show how preferred alternatives are determined based on the fuzzy utilities of the alternatives using fuzzy ranking. Section 7 presents the calibration approach. Section 8 shows comparative results of application of the model and binomial logit models to real-life data collected in the Netherlands. Finally, the concluding section 9 reviews the results presented in this paper, and outlines future research directions.
2 Modeling Multi-Modal Travel Choice Behavior

In this section, we briefly describe previous\(^1\) modeling approaches (random utility models, production rule systems) for travel choice behavior in public transport networks. Moreover, we will motivate the use of a fuzzy logic approach. That is, we will discuss why fuzzy sets may better describe humans' representation and appraisal of both trip chain attributes as well as the hypothetical utility of a trip chain alternative. Moreover, we argue that fuzzy inference can be used for improved description of the human decision process that is largely based on experience and habit.

2.1 Traditional modeling approaches

Individual trip choice models have been considered by a large number of researchers worldwide. For an elaborate overview, see [1]. These models serve several purposes, namely:

- Operationalisation of trip choice theory by specification of quantifiable relations between variables.
- Estimation of parameters and testing the aforementioned relations.
- Testing the hypothesized theories using observations.
- Estimation choice behavior in new situations.

Many types of models have been proposed, most of which are either 'random utility'-type models or production rule systems. Random utility models aim to predict the travel choice probabilities from a discrete set of alternatives. These models are probabilistic, implying differences in trip choice behavior among individual travelers. An entirely different type of model is the production rule system (see [2]). In these systems, rules, procedures, options and relationships, codified using systematic methods, represent the knowledge. This approach has shown potential for modeling complicated cognitive processes.

2.2 Motivation for using fuzzy logic and fuzzy sets

In this paper we propose a fuzzy based modeling approach to describe travelers' choice behavior in urban public transport networks. Basically, the motivation for using fuzzy sets and fuzzy reasoning is based on three arguments, namely:

1. Inherent vagueness in human perception and appreciation of trip chain attributes.
2. Knowledge and experience-based structure of the human decision process using a limited set of decision rules and approximate reasoning to decide in unknown situations.

\(^1\) In the literature, the term route choice is sometimes used, since trip choice implies a route in a public transport network. However, since also the modes, and the departure time are implied by the decision of the traveler, we refrain from the term route choice and use travel choice instead.
3. Non-linear, partially compensational, and partially lexicographic character of the (combined) appraisal of trip chain attributes; flexibility of fuzzy approach compared to random utility models.

When choosing a route, travelers are faced with uncertainty. On the one hand, travelers cannot predict the precise characteristics (attributes) of a route due to the randomness in these characteristics. In illustration, a traveler can not foresee whether or not a train will be delayed. We will refer to this type of uncertainty, describing (partially) unpredictable phenomena as randomness. Randomness in processes can be described by probability theory. With respect to travel choice modeling, random utility models convey this randomness (e.g. the in-vehicle time equals 10 minutes with probability 0.9, while with probability 0.1 the in-vehicle time equals 30 minutes).

However, uncertainty is not only related to randomness. Ambiguity relates non-random uncertainty with respect to fixed characteristics of an object or an event. It reflects among other things the imprecision of human perception and appraisal of characteristics of an object or an event. Clearly, we can assess whether the in-vehicle time is over 15 minutes. For example, if “long” is a set defined by trips longer or equal to 15 minutes, a computer would not recognize a trip of length 14.99 minutes as being a member of the set “long”. Nevertheless, in natural language, humans will certainly perceive a trip with an in-vehicle time of 14.99 minutes as “long”, due to the ambiguity or vagueness of the predicate long.

Summarizing, randomness relates to the non-deterministic nature of the travel choice behavioral process, and can be modeled adequately using probabilistic approaches (e.g. random utility approach). Vagueness or ambiguity on the one hand relates to the way travelers perceive and appraise attributes. On the other hand, it accounts for the fact that they rarely have exact information. Although several attributes, such as access and egress time, travel time, etc., are exact, other attributes are inherently vague. Examples of these vague attributes are comfort and safety. The traditionally used choice models cannot incorporate vagueness. However, fuzzy sets can be used. These sets are able to capture meanings and interpretations of words expressed in linguistic terms. For a detailed account of the applicability of fuzzy set theory in traffic and transportation, see [6].

Current travel choice models have been criticized for not realistically modeling human decision-making mechanisms (see [3] and [4]). We argue that travel choices are mostly made by comparing the choice situation at hand with experienced situations, which are reflected in a (limited) set of fuzzy decision rules. Based on the extent, to which the choice situations match these rules, a travel choice decision is made. That is, if a situation exactly matches a previously experienced situation, a traveler will in general act the same. Otherwise, a traveler will combine his experiences (interpolation) to come to a travel decision. Fuzzy logic is very suitable to describe this approximate reasoning process. That is, we argue that the fuzzy production-rule-like approximate reasoning mechanism realistically represents human decision processes.
Finally, a fuzzy logic approach does not constrain the structure of the appraisal and decision processes. On the contrary, random utility models are compensatory, for instance implying that high in-vehicle times may be compensated by low waiting times at transfers, and vice versa. Consequently, lexicography in the travel choice process is neglected. However, since in the fuzzy approach we do not a priori specify which attributes are used in the different rules, the resulting model is principally both compensatory as well as lexicographic. In other words, while in one rule, influences of in-vehicle time and waiting time are described by a compensatory mechanism (e.g. by the rules IF \( t = \text{'short'} \) AND \( w = \text{'long'} \) THEN \( u = \text{'medium'} \) and IF \( t = \text{'long'} \) AND \( w = \text{'short'} \) THEN \( u = \text{'medium'} \)), other rules may consider in-vehicle times and waiting times separately (e.g. IF \( t = \text{'high'} \) THEN \( u = \text{'low'} \) and IF \( w = \text{'very high'} \) THEN \( u = \text{'low'} \)). In illustration, see also table 1.

Summarizing opposed to random utility models, using the fuzzy model proposed in this paper, we do not need to specify the structure of the model a priori. This significantly improves the flexibility of the modeling approach.

At this point, let us remark that besides representing the human decision process more realistically, using a limited set of rules yields a mathematically tractable model, that can be calibrated using real-life observations. Consequently, the common calibration problem in fuzzy approaches (see [6]) can be handled.

\[2\] Both the fuzzy attributes in-vehicle time \( t \), waiting time at transfer \( w \), the fuzzy utility \( u \), as well as the fuzzy decision rules are discussed in the remainder of the paper.
3 Outline of FLiRT-Model

In this section, we present an overview of the FLiRT (Fuzzy Limited rule base Travel choice) model presented in the remainder of this paper. The approach consists of four steps. For each individual \( p \) facing the choice between alternatives \( j \) in the choice-set \( C_p \) with attributes \( x_{kj} \), where \( k \) denotes the \( k \)-th attribute, the following steps are performed (see Figure 1):

1. Fuzzification of the scalar (crisp) attributes \( x_{kj} \) of the respective trip alternatives \( j \) (section 4).

2. Fuzzy inference using a rule base consisting of a limited number of rules (section 5):
   a. determining level-of-fulfillment, i.e. the extent to which the antecedent of the rules are true;
   b. determining of rule-firing, i.e. the level in which the consequence of a rule is activated;
   c. aggregation of consequences of the respective rules; yielding a fuzzy utility \( u_j \) of trip alternatives \( j \).

3. Defuzzification of the fuzzy utilities \( u_j \) yielding crisp utilities \( U_j \) for all alternatives \( j \) (section 6).

4. Comparing the defuzzified utilities \( U_j \) and determining the traveler's choice \( I \) (section 6).

In the fuzzification phase, the relevant trip chain attributes are represented by fuzzy sets, or rather the degree with which they belong to a fuzzy set (degree-of-fulfillment). This yields a fuzzy representation of attributes that we believe better describes human perception and appraisal. The fuzzified attributes serve as an input to the fuzzy inference phase. Here, the fuzzy rule base, consisting of a limited number of fuzzy decision rules relating fuzzy attributes to fuzzy trip chain utilities, is of dominant importance since it represents the knowledge and experience of the travelers. The fuzzy inference mechanism triggers rules exactly for each known situation, while for unknown situations, a trip chain utility is determined by approximate reasoning, which can roughly be considered as non-linear interpolation or extrapolation of known situations. The resulting fuzzy utility is consequently defuzzified, that is, represented as a scalar value. Comparing the defuzzified utilities of considered trip alternatives yields the traveler's choice.

In general, calibration of fuzzy models is complicated due to the large number of parameters (membership functions) and unknown model relations (IF-THEN rules). To this end, expert knowledge and 'trial and error' are commonly used (see [6]). These approaches fail when no real understanding of the modeled processes is available. Fortunately, since the FLiRT model uses only a limited set of rules, calibration of the rule base by optimizing the so-called %-right-statistic is possible. Suitable membership functions are determined by
engineering judgement, while a genetic algorithm is used to establish the rule base.

Figure 1 - Outline of the FLiRT model.
4 Fuzzy Set Description of Attributes and Utilities

In this section, we show how fuzzy sets can be used to describe trip chain attributes and utilities of trip chains, given the ambiguity of traveler’s perception. First, we will discuss which time-related attributes are considered in the ensuing of the paper. We will show how fuzzy sets can be used to describe these attributes and utilities.

4.1 Modeling trip attributes using fuzzy sets

In this section, we introduce the concept of a fuzzy set and show how fuzzy sets can be used to represent the trip chain attributes of distinct alternatives. For convenience, we only consider time-related attributes:

1. Total in-vehicle time \( t \) (total time spent in the various transport modes used).
2. Access time \( a \) (time spent walking from the origin to the access point of the public transport system).
3. Egress-time \( e \) (time spent walking to the destination).
4. Waiting time \( w_0 \) at the first stop (time spent waiting for the first transport service in the trip chain).
5. Total waiting time \( w \) at all transfer points (total average time spent waiting for the next transport service at transfer points).

For sake of simplicity, we have aggregated the mode-specific trip attributes. That is, we do not consider differences in traveler’s appraisal of distinct modes within the trip chain. Mechanisms to incorporate mode-specific attributes are proposed in [7].

4.2 Crisp sets versus fuzzy sets

In traditional logic, a (crisp) set is a collection of elements that share a specific property, or not. For instance, we can define the set \( T_{(0,10)} = \{ t \mid 0 \leq t < 10 \} \) by all in-vehicle times \( t \) smaller that 10 minutes. Clearly, any in-vehicle time \( t \) is either a member of \( T_{(0,10)} \) or not. In other words, \( t \) has either a membership \( \mu \) equal to one when \( t \in T_{(0,10)} \), or zero whenever \( t \not\in T_{(0,10)} \). In the latter case, \( t \) is said to be a member of the complement \( T^C_{(0,10)} \) of \( T_{(0,10)} \). Note that we can also use the so-called membership function \( \mu \) to identify the set \( T_{(0,10)} \) (i.e. \( \mu(t) = 1 \) for all \( 0 \leq t < 10 \) and \( \mu = 0 \) elsewhere). We can use these crisp sets to describe the appraisal of travelers of specific attributes in terms of linguistic sets. For instance, we can define the set of 'low' in-vehicle times by the set \( T_{(0,10)} \) (see figure 2).
However, the use of crisp sets to describe the way in which humans perceive and appraise trip attributes such as in-vehicle times, waiting times, access and egress times, is limited since the thresholds defining these sets are vague and imprecise (see section 2.2). For one, the human mind is unable to (operationally) represent a real number exactly: ask ten people to indicate when 10 minutes have passed and each individual will indicate a different period length. In other words, the operational representation of exact real numbers is ambiguous (erroneous representation-of-real-numbers argument). Secondly, it is unlikely that humans consider in-vehicle times of 9.9 minutes as low, while an in-vehicle time of 10.1 minutes is considered high. In other words, representation of descriptive sets is mostly vague (vague appraisal argument).

To improve description of human-mind representation of sets describing specific properties, the concept of fuzzy sets is very useful. Similar to a crisp set, we can define a fuzzy set by its membership. The membership function \( \mu \) of a fuzzy set typically has values in the range \([0, 1]\) (although this is not necessary), opposed to crisp sets having membership values that are either 0 or 1. The independent variable in-vehicle time is the so-called base-variable of the fuzzy linguistic set. In illustration, the fuzzy set of low in-vehicle times can be defined by the membership function depicted in figure 2.

In addition to the improved description of vague and erroneous perception, representation, and appraisal, fuzzy sets can also be used to describe inherently vague attributes such as safety and comfort. In this case, the scale of the base-variable is undetermined and can be arbitrarily chosen.

### 4.3 Fuzzy trip chain attributes

In this paper we represent trip chain attributes by fuzzy linguistic sets. The term set \( S_k \) consists of linguistic terms \( T_k \) such as 'low', 'medium' and 'high' indicating the linguistic sets for attribute \( k \) for \( i = 1,\ldots, L_k \). Note that the term sets are defined for each attribute \( k = 1,\ldots, K \). The membership functions \( \mu_k \) used are bell-shaped, and completely determined by their mean \( \alpha_k \) and deviation \( \sigma_k \) for the different terms \( T_k \):

\[
\mu_k(x | \alpha_k, \sigma_k) = \exp\left(-\frac{1}{2}\left(\frac{x - \alpha_k}{\sigma_k}\right)^2\right)
\]  

(1)
The memberships $\mu_k^i$ specifying the 'lowest' and the 'highest' linguistic variables (usually $T_k^1$ and $T_k^i$) are equal to 1 in the regions $(-\infty, \alpha_k^i]$ and $[\alpha_k^i, \infty)$ respectively. Figure 3 shows the linguistic variables used in this paper, and the membership functions that fuzzify the time-related attributes.

Finally, for each of the attributes $k$, the term set $S_k$ is extended with an additional linguistic term, namely $T_{k+1} = 'all'$ having a membership value equal to 1 on the entire real axis. As will become clear in the remainder of this paper, term 'all' is used to reflect travelers indifference for a specific attribute.

4.4 Fuzzification of trip chain attributes

Having defined the fuzzy trip chain attributes, the (objectively measured) crisp trip chain attribute values can be fuzzified, that is, represented as linguistic sets. This is done using the membership functions $\mu_k^i$ of attribute $k$. These membership functions reflect the extent to which a crisp attribute value belongs to the different linguistic terms. In illustration, let the in-vehicle time term-set be given by $\{\text{'low', 'moderate', 'high'}\} = \{T_k^1, T_k^2, T_k^3\}$. The linguistic terms $T_k^i$ reflect the fuzzy sets defined by the membership functions $\mu_k^i$. For a specific alternative $j$ with in-vehicle time $t_j$ the $t_j$ is represented by the set of pairs $t_j = \{T_k^i, \mu_k^i(t_j)\}$. More generally, given the crisp attribute value $x_{kj}$ of the $k$-th attribute of alternative $j$, the fuzzified attribute $x_{kj}$ is defined by the collection of pairs:

$$x_{kj} = \{T_k^i, \mu_k^i(x_{kj})\}$$  \hspace{1cm} (2)

4.5 Fuzzy utilities

Similar to the linguistic terms $T_k^i$ representing trip chain attribute $k$, we will also use linguistic terms to specify the fuzzy utility of an alternative. By convention, the term-set $S_{k+1}$ represents the terms $T_{K+1}$ of the fuzzy utility for $i = 1, ..., K+1$. The fuzzy utility or preference is determined based on the fuzzified trip chain attributes $1 \leq k \leq K$ of the alternative using approximate reasoning described in the following section. This generally results in $I_{K+1}$ pairs $u_{kj} = \{T_{K+1}^i, \mu_{K+1}^i\}$. Note that only relative locations of membership functions $\mu_{K+1}^i$ are decisive (i.e. the scale of the fuzzy utility is ordinal rather than cardinal).
5 Approximate Reasoning using FLiRT-Model

This section outlines the notion of rule bases and fuzzy inference, and discusses the concept of approximate reasoning. Moreover, we discuss the justification for the choice of the rule base structure in the FLiRT model.

5.1 Fuzzy decision rules

Let us recall that in fuzzy logic, knowledge is stored using **fuzzy decision rules** having the following general form:

\[ \text{IF antecedent proposition THEN consequent proposition} \]  

Here, the **antecedent proposition** reflects the extent to which fuzzified values \( x_{kj} \) of the \( k \)-th attribute of trip chain alternative \( j \) in the choice-set \( C_p \) of individual \( p \) are equal to the set of fuzzy linguistic terms \( T_{k}^{(r)} \) of rule \( r \), where the function \( i(r) \) maps the rule number \( r \) to the \( i(r) \)-th linguistic term in \( S_k \). The extent to which the terms \( T_{k}^{(r)} \) are true, are combined using **AND-operations**. In illustration, an example rule is:

\[ \text{IF } \hat{t}_j = \text{'low'} \text{ AND } \hat{a}_j = \text{'moderate'} \text{ THEN } \hat{y}_j = \text{'high'} \]

or, more generally:

\[ \text{IF } x_{ij} = T_{ij}^{(r)} \text{ AND } \ldots \text{ AND } x_{kj} = T_{kj}^{(r)} \text{ THEN } y_{K^r} = T_{k^r,j}^{(r)} \]  

Table 1 shows a subset of the calibrated rules in the FLiRT model.

At this point, let us remark that part of the flexibility of the approach (i.e. the fact that the model can suit partially lexicographic and partially compensational orderings) is accommodated by the term 'all'. In illustration, the following rules can be present in the rule base:

\[ \text{IF } \hat{t}_j = \text{'moderate'} \text{ AND } \hat{a}_j = \text{'high'} \text{ AND } \hat{w}_j = \text{'moderate'} \text{ THEN } \hat{y}_j = \text{'low'} \]

\[ \text{IF } \hat{t}_j = \text{'high'} \text{ AND } \hat{a}_j = \text{'moderate'} \text{ AND } \hat{w}_j = \text{'moderate'} \text{ THEN } \hat{y}_j = \text{'low'} \]

\[ \text{IF } \hat{t}_j = \text{'all'} \text{ AND } \hat{a}_j = \text{'all'} \text{ AND } \hat{w}_j = \text{'high'} \text{ THEN } \hat{y}_j = \text{'low'} \]

Rules (6) and (7) illustrate how compensation works in a fuzzy approach. These rules show that a moderate in-vehicle time compensates for a high egress time and vice versa. Rule (8) shows how a lexicographic mechanism works: irrespective of the in-vehicle time and the egress time, the utility of an alternative is low whenever the waiting time is high. In other words, the high waiting time cannot be compensated for by, for example, low in-vehicle times or low egress times.
5.2 Fuzzy inference and approximate reasoning

Similar to traditional logic, fuzzy logic stipulates that if the antecedent proposition is true, so is the consequent proposition. However, in fuzzy-set terms, the antecedent proposition can also be partially true, yielding a partially true consequent proposition.

The level-of-fulfillment determines the degree to which the antecedent is true. In this paper, the level-of-fulfillment of the antecedent part of rule \( r \) equals the membership \( \mu_{K^r_j}(x_{kj}) \) of the observed \( k \)-th attribute value \( x_{kj} \) of alternative \( j \).

To combine the different attributes, the AND-operator is operationalized (see among others [13]) in the so-called inference-step. The FLIRT model uses the minimum operator to establish the gross level-of-fulfillment \( \phi^r_j \) of rule \( r \) for alternative \( j \). In illustration, consider rule (4). Suppose that for in-vehicle time \( t_i \) we have \( \mu_{\text{low}}^r(t_i) = 0.5 \), while for the egress time \( e_i \) we have \( \mu_{\text{moderate}}^r(e_i) = 0.8 \). The level-of-fulfillment of rule (4) equals \( \phi^r_j = \min\{0.5, 0.8\} = 0.5 \). Note that when the antecedent of attribute \( k \) equals 'all', this attribute becomes irrelevant in rule \( r \). In other words, the level-of-fulfillment \( \phi^r_j \) is independent of the value of the \( k \)-th attribute. The level-of-fulfillment \( \phi^r_j \) determines the so-called firing of rule \( r \), describing the degree to which the antecedent part of rule \( r \) is true, and consequently also the extent to which the consequent part holds (just like in traditional logic). In other words, the consequent proposition is true to the degree \( \phi^r_j \). For each rule \( r \) the resulting consequent part is defined by the pair \( \{T^r_{K+1}, \phi^r_j\} \). Considering rule (4), the fuzzy consequence for rule \( r \) is \( u_j = \{\text{high}, 0.5\} \).

In the final step, the consequences are aggregated using the maximum operator. That is, for a specific alternative, the firing of each rule \( r \) is determined, and the consequent fuzzy utilities of each separate rule are combined into a single fuzzy utility. More precisely, for each term \( T^r_{K+1} \) in the fuzzy utility term set \( S_{K+1} \), the membership of the gross fuzzy utility \( u_j \) of alternative \( j \) equals:

\[
\mu^r_{K+1,j} = \max_{r \text{ subject to } |r| = 1} \phi^r_j
\]

The fuzzy utility \( u_j \) is consequently defined by the collection of pairs \( \{T^r_{K+1}, \mu^r_{K+1,j}\} \). This implies that the (gross) membership \( \mu^r_{K+1,j} \) of the aggregate fuzzy utility is defined by:

\[
\mu^r_{K+1,j}(u) = \max_{i=1,...,K+1} \mu^r_{K+1,j}(u)
\]

for all \( u \).

5.3 Rule base choice

In some fuzzy logic approaches, a so-called full rule base is used. This means that for each possible combination of terms \( T^r_k \) of the \( k \)-th attributes, a rule is established. For each of these rules, the consequent proposition needs to be determined. That is, for each rule \( r \) the linguistic utility term \( T^r_k \) is set. Usually, this is done using both expert knowledge of the underlying decision process as well as trial-and-error. Depending on the number of linguistic terms in the
respective term sets $S_k$, the number of rules $N$ in a full rule base, with $N$ defined as:

$$N = \prod_{k=1}^{K} (l_k + 1)$$

(11)

can be considered. In illustration, considering five attributes each having four linguistic terms (excluding 'indifferent') yields a rule base of $5^5 = 3125$ rules. It is unlikely that humans will (and will be able to) store and use this amount of rules.

As a consequence, we have assumed that rule bases consist of a limited number of rules. Each rule $r$ is represented by a vector $i_r = (i_1, i_2, \ldots, i_{K+1})$ indicating both the attribute terms $T_k$ for $1 \leq k \leq K$ determining the antecedent propositions, as well as the utility terms $T_{K+1}$ determining the consequent proposition. In the operational model, the vector $i_r$ is represented uniquely by a single integer value $z_r$, which is achieved by the following mapping:

$$z_r = \sum_{k=1}^{K+1} (i_k \cdot N_k) \text{ with } N_k = \sum_{K+1}^{K} i_k$$

(12)

Vice versa, the elements $i_k$ of $i_r$ can be determined from $z_r$. Thus, a rule base $r$ of $M$ rules can be efficiently represented by a vector of $M$ integer numbers $z_r$. 
6 Choice of trip alternative

Having determined the fuzzy utility $u_i$ for each of the alternatives $j$, we need to compare these fuzzy utilities to establish which alternative $j^*$ is chosen by individual $p$. Three approaches have been considered:

- Defuzzification the fuzzy utility $u_i$ and choose the alternative $j^*$ with the highest defuzzified utility $U_j$.
- Comparison of the fuzzy utilities $u_i$ using for instance fuzzy ranking, yielding the best alternative $j^*$.
- Direct comparison of fuzzy utilities $u_j$ by means of the convolution principle.

Let us now discuss these different methods.

6.1 Defuzzification and comparison of crisp utilities

In [8] defuzzification and consequent crisp ranking is used to establish travelers’ choices. That is, the authors defuzzify the fuzzy utilities using the center of gravity of the fuzzy utility:

$$U_j = \frac{\int u \cdot \mu_{K+1,j}(u)du}{\int \mu_{K+1,j}(u)du} \quad (13)$$

and subsequently choose the alternative with the highest defuzzified utility.

It appears that defuzzification ‘smoothes’ important properties of the fuzzy utilities, thereby having a negative impact on model calibration. Especially rules with low fuzzy costs appear counterintuitive, meaning that for example in-vehicle time = ‘high’ and transfer time = ‘moderate’ resulted in low fuzzy costs. This phenomenon can be explained from eq. (13). On the one hand, low fuzzy costs hardly influence the denominator of eq. (13), since fuzzy costs are represented on a [0,1]-interval. On the other hand, low fuzzy costs are present in the nominator of eq. (13), thereby scaling the denominator. In other words, low fuzzy costs are no more than a scaling factor of the higher fuzzy costs. As a result, the antecedent of rules with low fuzzy costs are of minor importance, which diminishes the interpretability of the rule base.

6.2 Fuzzy ranking

Fuzzy ranking does not have the smoothing effect. Another interesting aspect of fuzzy ranking is the distinction of traveler classes with different attitudes towards risk [5]. An alternative scheme is proposed by Kaufmann and Gupta [10]. Results of application of fuzzy ranking are reported in [9]. Nevertheless, the center of gravity method has been applied in this paper.
6.3 Direct comparison of fuzzy utilities

Recently, we have developed an alternative approach to compare fuzzy numbers. Contrary to the approaches discussed in the previous sections 6.1 and 6.2, the presented approach enables direct comparison of fuzzy utilities. That is, we determine the 'degree' $0 \leq \beta \leq 1$ to which a fuzzy number $\mu_j$ (i.e. the fuzzy utility) is larger than another fuzzy number $\mu_i$. Let $\mu_j$ and $\mu_i$ denote the membership functions of $\mu_j$ and $\mu_i$ respectively. In case on an binary choice (two alternatives: $j = 1,2$), we define $\beta_1$ by:

$$\beta_1 = \frac{1}{B_1 \cdot B_2} \int_{\mu_1} \mu_1(u) du dw \text{ with } B_j = \int \mu_j(u) du$$

We will refer to $\beta_1$ as the representative relative preference of alternative 1 over 2: when $\beta_1 > 0.5$ (i.e. $\mu_1$ is vaguely larger than $\mu_2$ to the degree $\beta_1$) the individual prefers alternative 1 to 2. As $\beta_1$ tends to one, the preference for 1 becomes 'less ambiguous'. Note that $\beta_2$, defined by the representative relative preference of alternative 2 over 1, satisfies: $\beta_2 = 1 - \beta_1$.

Note the similarity of this measure and the probability $\pi$ that one random number is larger than another. That is, if we consider random numbers $\tilde{u}_j$ and $\tilde{u}_2$, then the probability $\pi_1$ that $\tilde{u}_1$ is larger than $\tilde{u}_2$ equals:

$$\pi_1 = \int_{\tilde{u}_1} \int_{\tilde{u}_2} f_1(u) du dw$$

where $f_j$ denotes the probability density function of the random number $\tilde{u}_j$.

---

$^3$ Note that the function $f = \mu/B$ defines a probability density function.
7 Calibration of FLiRT-model

In this section we present the calibration approach, which is based on minimization of the so-called %-right-statistic. The procedure consists of two steps.

1. Determine suitable membership functions by engineering judgement.
2. Apply a Genetic Algorithm to establish the rule base based on optimization of the %-right statistic (see [8]). In this step, both the antecedent proposition and the consequent proposition of the limited set of rules are determined.

However, let us first discuss how we assess the performance of the model.

7.1 Performance measure

To assess the quality of the model, a performance measure is necessary. However, due to the nature of the data used for model calibration, only a few are applicable. One of the few sensible measures is the fraction of correctly predicted travel choices (the so-called %-right-statistic).

For each individual \( p \) with alternatives \( j \), we define an indicator \( 1_p(j) \) that is equal to 1 if the defuzzified utility \( U_j \) is both maximal (i.e. the model predicts that individual \( p \) chooses \( j \)) and actually chosen by \( p \). Let \( P \) be the total number of individuals in the sample, and let \( R \) be a specific rule base. Then, the %-right-statistic \( J(R) \) is defined by:

\[
J(R) = \frac{\text{100\%} \sum_{p=1}^{P} 1_p(j^*)}{P}
\]  

(16)

It has been argued by among others [1] that there are some drawbacks using this performance measure. For instance, a low %-right statistic does not necessarily imply poor model performance (although the contrary does hold, i.e. a high %-right statistic does imply good model performance), which is especially apparent in the case of repeated choice situations. However, in our case, each choice situation is unique. We have used the %-right-statistic in the remainder of this paper.

Recently, we have considered an alternative performance measure based on the representative relative preference \( \beta_j \) (eq. (14)). For a rule base \( J \), we have defined the following performance measure:

\[
J(R) = \frac{1}{P} \prod_{p=1}^{P} \prod_{j \in C_p} (\beta_j)^{1_p(j^*)}
\]  

(17)

\[\text{In this paper, we judge the model performance mainly by considering the predictive performance of the models. Analysis of the explanatory power of the models is beyond the scope of this paper.} \]
In this case, we aim to maximize the products of the representative relative preferences of the chosen alternatives. In rough terms, this measure can be seen as the fuzzy equivalent to the likelihood function used in estimation of random utility models (see [1]).

7.2 Choice of membership functions
We have determined the values for \( \alpha_k \) and \( \sigma_k \) specifying the membership functions by our own engineering judgement and trial-and-error. For example, we have chosen values that are expected to reasonably reflect low, moderate and high in-vehicle times. Moreover, we have assured ourselves that the whole observation range was covered by the resulting membership functions for all attribute values present in the sample.

Let us note that with respect to the membership functions, it appears that the model performance is not sensitive with respect to small changes in \( \alpha_k \) and \( \sigma_k \). Nevertheless, sensitivity analysis should provide definite answers with respect to this issue.

7.3 Calibration of FLiRT-model using a Genetic Algorithm
Having established the membership functions for all attributes and consequent propositions, the rules of the rule base are calibrated. Aim of the model calibration is to find the rule base \( R^* \) optimizing the %-right-statistic. Motivated by the large and discrete solution space, the irregular and non-linear optimization objective, and the non-uniqueness of the optimal solution, we have chosen to use a Genetic Algorithm (GA) to approximate the optimal solution.

GA's are based on the principles of evolution. That is, it describes how individuals in a generation evolve and reproduce. In the proposed GA calibration approach, individuals are rule bases. We have shown how each rule \( r \) is encoded by an integer \( z_r \) representing the vector \( l_r \) indicating both attribute terms, and the utility term. Consequently, a rule base of \( M \) rules can be encoded as an \( M \)-dimensional vector \( r \) of integers \( z_r \). A generation \( G_m \) is a collection of rule bases \( R; \) with \( 1 \leq n \leq N \), where \( M \) indicates the generation number. The initial generation \( G_0 \) is determined by randomly drawing \( M \) integer numbers reflecting arbitrary rules.

For each rule base \( R; \) in the \( M \)-th generation \( G_m \) we establish the so-called fitness \( f_n \) by determining the number of travelers in the data whose travel choice is correctly predicted. Based on the evolution principle, the fitness \( f_n \) of an 'individual' \( R; \) determines the probability that an individual will survive and reproduce. Survival implies that a rule base \( R; \) will still exist in generation \( G_{m+1} \). To ensure quick convergence of the GA-scheme, two rule bases having the best fitness survive.

Reproduction is based on the principles of selection and crossover. The so-called parent rule bases \( R_p \) and \( R_q \) are selected from generation \( G_m \) by chance. The probability that rule base \( R_p \) and \( R_q \) are selected depends on their fitness \( f_p \) and \( f_q \) respectively. A high fitness implies a high probability of being selected from generation \( G_m \). When two parents (rule bases \( R_p \) and \( R_q \)
have been selected, crossover (exchanging) of specific elements (rules $r_p$ and $r_q$) yields the so-called offspring rule bases $R_p'$ and $R_q'$. During crossover rules $r_p$ and $r_q$ of $R_p$ and $R_q$ are interchanged. We have used so-called parameterized uniform crossover to (randomly) select which elements are exchanged.

Finally, some elements $r'$ of the offspring $R'$ are selected for mutation by chance. The probability of mutation depends on the firing $\phi'$ of rule $r'$: when $\phi'$ is small, the probability that $r'$ is selected for mutation is high. On the contrary, when $\phi'$ is large, $r'$ is selected for mutation with a small probability. When an rule $r'$ is mutated, a new rule is determined by drawing an integer $\zeta$ (representing a rule $r$) from a uniform distribution, and replacing $z'$ by $\zeta$. Again, we draw new integers until the respective rules are consistent with the other rules in the rule base.

The process of selection, crossover, and mutation is repeated until generation $G_{m+1}$ consists of $N$ rule bases. The approach is consecutively applied to the new generation $G_{m+1}$, and repeated until it is unlikely that the fitness values will improve from one generation to another. For a more detailed account on GA, we refer to [9].
8 Application of the FLiRT-model

In this section, we show the results of application of the FLiRT model. To this end, we discuss the calibration results achieved by considering revealed preference data collected for public transport networks in major Dutch cities. Moreover, the model is verified using a control set, and compared with a binomial logit model (see [11] and [12]).

8.1 Data-description

In 1986 and 1987 a large survey was conducted in major Dutch cities in order to establish the qualitative appraisal of time-related trip chain attributes such as in-vehicle time, access and egress time, waiting time at first stop and waiting time at transfers (see [11]). Twenty-five so-called origin areas (situated at the city suburbs) and matching destination areas (located in the city center) were selected. The survey was conducted by interviewers at the initial stops in the origin areas. The selection of origin-destination pairs was based on the availability of at least two public transport alternatives from origin to destination.

The collected data contains mode-specific travel time components of trips in a urban public transport network. Furthermore, information on the transfers, such as frequency of the subsequent transport service in the chain, walking times in the transfer points, waiting and transfer comfort, is available. Information on the actual waiting times, that is the time between arrival of one transport service and the departure of the subsequent transport service in the chain is not available. Instead, the average waiting time at transfer points was approximated by half the average headway (inverse frequency) of the public transport service, while the waiting time at the first stop was approximated from the frequency using a Weber function. For more details, we refer to [11].

8.2 Experimental set-up

In total, 1095 individuals who have two trip alternatives were surveyed. Of this sample, 400 individuals were randomly selected for model calibration purposes (calibration set). The remaining set of 695 individuals was used to validate the model (control set). For the considered test cases, the attributes used in the FLiRT model are discussed in the previous section. The membership functions describing both the attributes as well as the utilities are depicted in figure 3. Initially, five membership functions (including 'all') were used for access time and egress time, while only three membership functions where used for total in-vehicle time, total transfer time and waiting time at the first stop. In the renewed model we use six membership functions for each attribute. Again these membership functions are based on expert knowledge. Furthermore, the amount of overlap between membership functions is reduced.

With respect to specification of the FLiRT-model, we considered 100 fuzzy decision rules. Applying the genetic calibration approach discussed in the previous section, the model was calibrated using the calibration set and tested on the control set. This was also done using a binomial logit model. Additionally,
the best performing logit model determined in has also been considered. For a description of the logit model calibration approach, we refer to [11]. In the remainder of this section, we discuss the calibration and verification results, and compare the model performance of the different models.

8.3 Example rule and firing mechanism

Before discussing and comparing the model performance, let us illustrate the workings of the FLIRT model by considering a set of rules (rows (1)-(10) of table 1) for an arbitrary individual $p$ in the calibration data set facing two alternatives. Table 1 shows the rules that are fired (column (g) of table 1). For this particular individual, for each of the linguistic terms reflecting the fuzzy costs (column (f)), given the linguistic terms in the antecedent part (column (a)-(e)).
From table 1 we observe that \( p \) considers the cost of alternative 1 as 0.85 'low' (row (1)) and 0.47 'very low' (row (2)). Alternatively, \( p \) considers the cost of alternative 2 as 1.00 'low' (row (6)), but only 0.05 'very low' (row (7)). Thus roughly said, the cost of alternative 1 are lower compared to the cost of alternative 2, from a fuzzy perspective. Hence, the former alternative is chosen.

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>trip chain attributes in antecedent part</td>
<td>fuzzy cost</td>
<td>firing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very high</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>-</td>
<td>low</td>
<td>0.85 (1)</td>
</tr>
<tr>
<td>very high</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>-</td>
<td>moderate</td>
<td>very low</td>
</tr>
<tr>
<td>high</td>
<td>very high</td>
<td>low</td>
<td>low</td>
<td>-</td>
<td>moderate</td>
<td>very high</td>
</tr>
<tr>
<td>high</td>
<td>very high</td>
<td>moderate</td>
<td>-</td>
<td>-</td>
<td>very high</td>
<td>0.07 (4)</td>
</tr>
<tr>
<td>moderate</td>
<td>high</td>
<td>high</td>
<td>low</td>
<td>moderate</td>
<td>high</td>
<td>0.00 (5)</td>
</tr>
<tr>
<td>very high</td>
<td>low</td>
<td>low</td>
<td>-</td>
<td>-</td>
<td>low</td>
<td>1.00 (6)</td>
</tr>
<tr>
<td>very high</td>
<td>low</td>
<td>low</td>
<td>-</td>
<td>-</td>
<td>moderate</td>
<td>very low</td>
</tr>
<tr>
<td>high</td>
<td>very high</td>
<td>low</td>
<td>low</td>
<td>-</td>
<td>moderate</td>
<td>0.04 (8)</td>
</tr>
<tr>
<td>high</td>
<td>very high</td>
<td>moderate</td>
<td>-</td>
<td>-</td>
<td>very high</td>
<td>0.04 (9)</td>
</tr>
<tr>
<td>moderate</td>
<td>high</td>
<td>high</td>
<td>low</td>
<td>moderate</td>
<td>high</td>
<td>0.00 (10)</td>
</tr>
</tbody>
</table>

Table 1 – Fuzzy rules and consequent firing for individual choosing alternative 1 over 2. The dashes indicate the term ‘all’.

8.4 Analysis of calibrated rule base

Having calibrated the FLiRT-model using the 400 observations in the calibration data set, it was found that of the rules in the rule base, 9 rules were not fired \( \phi = 0 \) in any choice situation. Consequently, for this particular case, 91 rules would have sufficed to attain the same model performance. The performance decrease when considering even less rules has not yet been studied.

The rules in the rule base are interpretable (see e.g. table 1). Moreover, by consideration of the calibrated rule base, it appeared that travelers where either indifferent with respect to waiting time at the first stop \( w_0 \) (reflected by the linguistic variable ‘indifferent’), or rules with \( w_0 \) not equal to ‘indifferent’ have very low firing values \( \phi \), implying a very small contribution of the fuzzy utility of these rules to the aggregate fuzzy utility of the alternatives\(^5\), which is in agreement with the results of [12].

8.5 Comparison %-right FLiRT-model and logit model

Having calibrated the fuzzy model using the genetic algorithm, we have observed that the predictive performance of the model improves significantly compared to the random utility models of [11]. On the calibration data set, the model established the preferred alternative correctly in 84% of the 400 choice

---

\(^5\) To check whether \( w_0 \) has a small influence on the choice process, we have recalibrated the FLiRT model without \( w_0 \). It was found that the predictive performance only decreased with 1%.

IRAIL Research School, December 1999

21
A calibrated logit model using the same trip chain attributes, namely $a$, $e$, $t$, $w$, and $w_0$, only determined the correct preferred alternative in 72% of all cases\(^6\). This holds equally for the control set: the FLiRT model correctly establishes the preferred alternative in 84% of all cases, while the logit model only determines the correct alternative in 73% of all instances. Concluding, the FLiRT model outperforms the logit model indisputably.

It can be argued that part of the improvements achieved are due to the fact that the %-right statistic is used to calibrate the FLiRT model. However, recent results achieved by using the alternative performance measure eq. (17) to calibrate the model (comparable to the Maximum-Likelihood method used for the logit approach), yielded the same model performance in terms of %-right predicted as when using the %-right statistic. Consequently, we are confident that the performance improvements are not caused by the considered calibration criterion.

---

\(^6\) Note that since only situations where two alternatives are present are considered, a random selection of the preferred alternative yields a predictive performance of 50%. In this case, 50% should be used as a benchmark, rather than 0%.

\(^7\) We also have considered the model of [11] yielding the higher %-right statistic, yielding 75% correctly predicted travel choices. In this case, also the attributes number of transfers and the walking time at transfer points were included in the model.
9 Closing

In this paper, we have established a fuzzy travel choice model FLiRT for the description of travel choice behavior in public transport networks. The fuzzy approach is justified by the vague and erroneous way travelers perceive and appraise the attributes of trip chains. Moreover, we believe that using a limited set of fuzzy decision rules better approximates traveler's choice behavior, that is largely based on knowledge, experience, and approximate reasoning-like mechanisms.

The paper also discusses calibration of the FLiRT model, which is a common problem in fuzzy modeling. However, opposed to most fuzzy models, the structure of the considered rule bases enables calibration by optimization. Basically, this is performed in two phases. In the first phase, the membership functions linguistically representing both the attribute values, as well as the utilities of the alternatives, are established. This is mainly done by expert knowledge. In the second phase of model calibration, the rule base is determined by application of a genetic algorithm. That is, from the set of admissible fuzzy decision rules, the subset of rules that optimize the percentage of correctly predicted choice situations (the %-right statistic) is determined.

Finally, we have briefly discussed the results of applying the calibrated fuzzy travel choice model to revealed preference data collected in public transport networks in major Dutch cities. Generally, the resulting rules in the rule base are interpretable. Based on the %-right statistic, we may conclude that the model performs very well: on the calibration set, 84% of the situations is correctly predicted, while in the control set, 84% is correctly predicted. This is a significant increase with respect to traditional binomial logit-type models, the performance of which is within the range of 72%-75% (depending on the model specification and the attributes taken into consideration) correctly predicted choice situations, on both the calibration set, as well as the control set. Based on this result, we argue that the human choice process is indeed better described using the proposed fuzzy approach.

9.1 Future research direction

To improve the model, several potentially beneficial research directions can be identified:

- **Sensitivity analysis.** Analysis of the performance of the FLiRT with respect to its model parameters (e.g. membership function parameters, number of rules in the rule base).
- **Application alternative methods to establish the preferred alternatives instead of simple defuzzification** (see [7]). Among other things, this would enable determining sensitivity of the probability that an alternative is chosen with respect to small changes in attribute values. This (indirectly) enables analyzing the relative weights of the distinct travel-time components.
• Development of calibration techniques to simultaneously calibrate both rules as well as membership functions.

• Assessment of potential to apply rule-chaining to incorporate the hypothesized hierarchical structure of the travel choice process (see [7]) similar to e.g. a nested-logit approach. By doing so, a mechanism to correctly aggregate mode-specific attributes (e.g. in-bus or in-train travel time) into the modeling approach can be established.

• Consideration of differences in attributes, rather than the absolute attribute-values. In doing so, the calibrated rule base will not contain counterintuitive rules. Moreover, the application area of the model will probably increase, since situations with very different travel alternatives can also be determined.

• Application of the approach to decision processes in alternative travel systems, for example simultaneous mode-, departure-, and route-choice behavior in interurban and urban traffic networks.

Given the significant performance improvements compared to traditional modeling approaches clearly shows the potential of the FLIRT model, justifying additional research efforts regarding further model improvements.

9.2 Acknowledgement

This publication is a product of the research programme Seamless Multi-Modal Mobility, carried out within the Netherlands TRAIL Research School for Transport, Infrastructure and Logistics, and financed by the Delft University of Technology.
References


Influence of Delays on Departure Times in Air Traffic

TRAIL Research School, Delft, December 1999

Authors:
Ir. G. Soto y Koelemeijer
Information Technology and Systems, Delft University of Technology
Ir. Hanno van der Steen
Management Services (AMS/QO), KLM Royal Dutch Airlines

Thesis supervisors:
Prof.dr. Geert-Jan Olsder
Information Technology and Systems, Delft University of Technology
dr. J.W. van der Woude
Information Technology and Systems, Delft University of Technology
Influence of Delays on Departure Times in Air Traffic

THRII Research School Delhi December 1990

Author

In: E. Sava, M. M. Cremers

Introduction: In the ICAO Annex 4 it is recommended to implement

Traffic Separation Procedures

ICAO procedures (DCA - 360) for Air Traffic Control

Air Traffic Flow Management (ATFM) and its Integration in

the ICAO, Joint Work

Identification of these procedures and implementation of them
Contents

Abstract

1 Introduction 1

2 The max-plus algebra 2

3 Structure of KLM's schedule 4

4 The KLM model 5

5 Extensions of the model 8

6 Conclusions 10

References 11

7 Appendix A 12

8 Appendix B 14
Abstract

We study robustness of timetables in air traffic. In particular, we analyse the influence of delays on departure times of flights. In order to do so, we consider different types of resources, such as aircraft, cockpit crew and cabin crew, which have to be synchronized in order to constitute 'flight'. Our model is capable of predicting the effect of a delay of one of the resources, e.g. a delayed cabin crew, on the departure times of all flights of a given timetable, without interference of a traffic controller. To obtain the model we use the max-plus algebra, which is a new tool for modeling transportation networks.
1 Introduction

Transportation networks are a typical example of what we call DESs (Discrete Event Systems). Examples of discrete events in a transportation network are departures and arrivals. For simplicity we shall only use departures. There are many modeling and analysis techniques for DESs, such as queueing theory, automata theory, max-plus algebra, Petri nets, generalized semi-Markov processes, perturbation analysis and so on. Every single one has its advantages and disadvantages. In this paper we concentrate on the max-plus algebraic method; there are well-developed tools available for analysis. For more about the max-plus algebra the reader is referred to [1].

In [2] a max-plus model is derived for the Dutch Railway system. An improved version can be found in [4]. These models are dynamic models because of the hourly pattern of the trains. Our goal is to present a static model to describe transportation networks. As an example we made a model from the KLM with real data. The reason that this is a static max-plus model is that every plane that departures has a unique flight number.

The paper is organized as follows. In section we give a motivation for the max-plus algebra. In section 3 we give the structure of the KLM’s schedule. The model is given in section 4 and an extension of the model is given in section 5. Finally we give a short conclusion in section 6.
In this section we give a motivation for the max-plus algebra concerning trans­portation networks. Let us consider Figure 1. There are four flights, indicated by 1, 2, 3 and 4 and thus four airplanes. The flying times are also given. Suppose now that the airplane with flight number 3 can not leave before the airplanes with flight numbers 1 and 2 have arrived. Let \( x_i \) denote the departure time of an airplane with flight number \( i \), then we are interested in the value of \( x_3 \). It is clear that the airplane with flight number 1 lands at \( x_1 + 5 \) and that the airplane with flight number 2 lands at \( x_2 + 7 \). Hence,

\[
x_3 = \max(x_1 + 5, x_2 + 7).
\]

In this equation we have two operators, namely \( \max \) and \(+\). This is a simple motivation for the \( (\max, +) \)-algebra. For simplicity, we write \( \oplus \) for \( \max \) and \( \otimes \) for \(+\), then (1) translates into

\[
x_3 = (x_1 \otimes 5) \oplus (x_2 \otimes 7).
\]

So, the max-plus algebra is defined by the operations addition \( (\otimes) \) and maximization \( (\oplus) \) applied to the real numbers, extended with minus infinity. Let us give some examples:

**Example 2.1**

1. \( 5 \oplus 3 (= \max(5, 3)) = 5 \)
2. \( 6 \otimes 8 (= 6 + 8) = 14 \)
3. \( 3^{3\oplus} (= 3 \otimes 3 \otimes 3 = 3 + 3 + 3) = 3 \cdot 3 \)
4. \( 4 \otimes 6 \oplus 5 (= \max(4 + 6, 5)) = \max(10, 5) = 10 \)

The unit element is \( e \) and has numerical value 0 and the neutral element with respect to \( \oplus \) is \( e \) and has numerical value \(-\infty\).
Example 2.2

1. $5 \otimes \varepsilon(= 5 + -\infty) = -\infty$
2. $5 \oplus \varepsilon(= \max(5, -\infty)) = 5$
3. $5 \otimes e(= 5 + 0) = 5$
4. $5 \oplus e(= \max(5, 0)) = 5$

In the sequel we will frequently see equations of the following kind:

$$x = A \otimes x \oplus d. \quad (3)$$

Such an equation can be solved (if the solution exists) by starting with, for example $x = 0$, where 0 is the zero vector, and by iterating equation (1) until $x$ after the $(k + 1)$-st iteration is equal to $x$ after the $k$-th iteration.
3 Structure of KLM's schedule

In this section we give the information, which will be needed to build a mathematical model.

There are three resources, namely 1, 2 and 3. Resource 1 denotes aircrafts, resource 2 denotes cockpit crew and resource 3 denotes cabin crew. Each flight needs at least one aircraft, a cockpit crew and a cabin crew and can not depart before all these resources are available. Because, at airports, these resources have to transfer from flight to flight, connections arise. These connections can easily be visualised (see Figure 2). Let us clarify Figure 2. The numbers beneath the vertical bars denote the flight numbers. The numbers above the vertical bars denote departure time (left) and arrival time (right). The lines between the vertical bars denote connections from flight with flight number $j$ to flight with flight number $i$ for aircrafts (upper), for cockpit crew (middle), and cabin crew (lower).

When an aircraft arrives at Schiphol, it is clear that this aircraft can not depart immediately; the minimal time that the aircraft needs before departing again is 50 minutes. The cockpit crew needs also 50 minutes before departing again after an arrival and the cabin crew needs 70 minutes. We will call these values ground time. In practice, the difference between the $(k+1)$-st departure time and the $k$-th arrival time for each resource separately is bigger than these values. The extra time $(\text{departure time} - (\text{arrival time} + (50 \text{ or } 70)))$ is called buffer time and is used to decrease or even eliminate possible delays. For airports, other than Schiphol, we assume that the time, needed to depart again after arrival, is 35 minutes for each resource.
Figure 2: Example of a network

Table 1: Renumbered flights for our example

<table>
<thead>
<tr>
<th></th>
<th>1004</th>
<th>9</th>
<th>1760</th>
<th>17</th>
<th>1033</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1553</td>
<td>10</td>
<td>1227</td>
<td>18</td>
<td>1034</td>
</tr>
<tr>
<td>3</td>
<td>1554</td>
<td>11</td>
<td>1228</td>
<td>19</td>
<td>1635</td>
</tr>
<tr>
<td>4</td>
<td>1333</td>
<td>12</td>
<td>1071</td>
<td>20</td>
<td>1909</td>
</tr>
<tr>
<td>5</td>
<td>1334</td>
<td>13</td>
<td>1072</td>
<td>21</td>
<td>1910</td>
</tr>
<tr>
<td>6</td>
<td>1873</td>
<td>14</td>
<td>1407</td>
<td>22</td>
<td>1367</td>
</tr>
<tr>
<td>7</td>
<td>1782</td>
<td>15</td>
<td>1711</td>
<td>23</td>
<td>1368</td>
</tr>
<tr>
<td>8</td>
<td>1501</td>
<td>16</td>
<td>1712</td>
<td>24</td>
<td>1018</td>
</tr>
</tbody>
</table>

4 The KLM model

In order to analyse the influence of delays on departure times at Schiphol airport we shall use the max-plus algebra. We are not interested in how delays are caused, only how they influence the future departure times. Therefore we need a mathematical model. We describe the model by means of an example, which is a subnetwork of the complete network. The subnetwork consists of 24 flights. Which flight number is connected to which flight number (i.e. some flights are not allowed to depart or cannot depart before some others have arrived) is given in the data (see appendix A). We can describe this as a max-plus model. First we renumber the flights (see Table 1).

The max-plus equations for this subnetwork are given by
\[
x_1 = d_1 \\
x_2 = \max(x_{24} + 180, x_1 + 180, d_2) \\
x_3 = \max(x_2 + 145, d_3) \\
x_4 = \max(x_3 + 170, d_4) \\
x_5 = \max(x_4 + 120, d_5) \\
x_6 = \max(x_5 + 135, d_6) \\
x_7 = d_7 \\
x_8 = \max(x_3 + 190, x_7 + 115, d_8) \\
x_9 = d_9 \\
x_{10} = \max(x_1 + 180, x_9 + 160, d_{10}) \\
x_{11} = \max(x_{10} + 120, d_{11}) \\
x_{12} = \max(x_{11} + 135, d_{12}) \\
x_{13} = \max(x_{12} + 105, d_{13}) \\
x_{14} = \max(x_{13} + 115, d_{14}) \\
x_{15} = \max(x_{11} + 135, d_{15}) \\
x_{16} = \max(x_{15} + 160, d_{16}) \\
x_{17} = \max(x_{11} + 155, d_{17}) \\
x_{18} = \max(x_{17} + 125, d_{18}) \\
x_{19} = \max(x_8 + 135, x_{18} + 145, d_{19}) \\
x_{20} = \max(x_1 + 200, d_{20}) \\
x_{21} = \max(x_{20} + 110, d_{21}) \\
x_{22} = \max(x_{21} + 120, d_{22}) \\
x_{23} = \max(x_{22} + 245, d_{23}) \\
x_{24} = d_{24}
\]

(4)

We clarify these equations by means of an example. Consider the equation for \(x_2\).

\[
x_2 = \max(x_{24} + 180, x_1 + 180, d_2).
\]

As one can see in Figure 2, the departure time of flight with flight number 2 depends on the arrival times of the flights with flight numbers 24 and 1. The traveling time of flight with flight number 24, which arrives at Schiphol, is 110 minutes and the connection with the flight with flight number 2 is a cabin crew connection. Hence, we add 70 minutes (ground time for cabin crew at Schiphol Airport) to the traveling time (110 minutes). The same can be done for the flight with flight number 1 (180=130+50).

The vector \(d\) is the scheduled departure time vector. The model can be written...
as

\[ x = A \otimes x \oplus d \]  

(5)

where \( A \) can be written as \( A^{ai} \oplus A^{co} \oplus A^{ca} \oplus I \), where \( A^{ai} \) contains information about the connection of aircrafts (i.e. traveling time plus ground time) etc. Remark that the solution of equation (3) is \( x = d \).

The symbol \( I \) denotes the unit matrix, where \( I_{ii} = 1 \) for all \( i = 1, \ldots, n \), and \( I_{ij} = 0 \) if \( j \neq i \). The model becomes

\[ x = A^{ai} \otimes x \oplus A^{co} \otimes x \oplus A^{ca} \otimes I \otimes x \oplus d = (A^{ai} \oplus A^{co} \oplus A^{ca} \oplus I) \otimes x \oplus d. \]  

(6)

The matrices \( A^{ai}, A^{co} \) and \( A^{ca} \) for our example can be found in Appendix B.

The influence of delays on actual departure times can now easily be derived. Change the appropriate elements of \( A^{ai}, A^{co} \) and \( A^{ca} \) to express the delay (for example if a plain with flight number \( j \) has a delay of 34 minutes and the flight number of the next flight is \( i \), then \( A^{aj}_{ij} = A^{aj}_{ij} + 34 \)) and we obtain \( A^{*ai}, A^{*co} \) and \( A^{*ca} \). Solve

\[ \bar{x} = (A^{*ai} \oplus A^{*co} \oplus A^{*ca} \oplus I) \otimes \bar{x} \oplus d. \]  

(7)

The influence of delays on departure times is now given by

\[ ddp = \bar{x} - d. \]  

(8)

**Example 4.1** If flight with flight number 1760 has a delay of 80 minutes and flight with flight number 1004 has a delay of 75 minutes, we get the following delays (see Figure 3). The boldface vertical bars in Figure 3 denote the delayed flights. The numbers on top of the vertical bars denote the delay of that particular flight.
5 Extensions of the model

As already mentioned, buffer times are needed to decrease or even eliminate possible delays. The more buffer times, the more stable is our timetable. On the other hand, from a commercial point of view, we want the buffer times to be as little as possible. For a good timetable, we need to find a balance. The max-plus algebra gives more insight in where and how well buffer times are distributed, given the timetables and the connections of the resources. The max-plus algebra is also a good tool to test how vulnerable the timetable is if delays occur.

It is interesting to know, which nodes are affected by a certain delay at a certain node. Therefore we introduce the following definition [3].

**Definition 5.1** \( M_{i,j} \) is the maximum delay of node \( j \) that does not cause a delay at node \( i \).

We obtain the \( M \) matrix by calculating

\[
M_{ij} = d_i - d_j - A_{ij}^*,
\]

where \( A_{ij}^* = \Theta_{k=1}^n(A^k)_{ij} \). Because a resource is used at most 8 times a day we say that \( n = 8 \). (Normally \( n \) represents the size of a square matrix.) If now a flight with flight number \( j \) has a delay of 75 minutes, then flight with flight number \( i \) has a delay of

\[
\max(75 - M_{ij}, 0).
\]

Figure 3: Delays
The matrix $M$ for our example can be found in Appendix B.

If a flight has a delay, the question arises, which resource causes the delay. Suppose that flight with flight number $i$ cannot depart before the flights with flight numbers $j$, $k$ and $l$ have arrived and $i$ has a delay. The connection $j \rightarrow i$ means that there is a aircraft connection, the connection $k \rightarrow i$ means that there is a cockpit crew connection, and the connection $l \rightarrow i$ means that there is a cabin crew connection. If we determine

$$C_{ij}^{ai} = d_i - d_j - A_{ij}^{ai}$$
$$C_{ij}^{co} = d_i - d_j - A_{ij}^{co}$$
$$C_{ij}^{ca} = d_i - d_j - A_{ij}^{ca}$$

and if all delays are known, one could easily find the resource which causes the delay. Remark that $C$ is defined as the maximum delay of node $j$ that not quite reaches the following node in the network, which is node $i$. This can be of interest, because if a aircraft and a cockpit crew have to wait for a cabin crew, which has a delay of 60 minutes one could think of adding a cabin crew, such that the flight is able to depart as scheduled.

To offer a good service, one could think of also model passenger flows. This can easily be obtained by adding some equations to the model. If in our example we want passengers from flight with flight number 9 to have a connection with flight with flight number 18 and if 35 minutes are needed to transfer we get

$$x_{18} = \max(x_{17} + 125, x_9 + 145, d_{18}).$$

Remark that the traveling time of flight with flight number 9 was 110 minutes. A fourth matrix, $A^{pas}$ can be constructed.
Conclusions

In this paper we presented a static max-plus model for transportation networks. It turns out that the influence of delays on departure times can be calculated in advance, like in the dynamic case. This can be used to compare different timetables. Because of the many delays on an arbitrary day, flights are switched and therefore it is impossible to verify if our model for the influence of delays is realistic. Further research includes the applications of the minmax-plus algebra, for the static case as well as the dynamic case.
References


Appendix A

In this Appendix we give the data, which is needed for the example in section 4, which consists of 24 flights. The meaning of the headings in Tables 2 is:

- Flnr: flight number
- Ind: index after renumbering
- ResAir: resource aircraft
- ResCoc: resource cockpit
- Rescab: resource cabin
- DepT: departure time
- ArT: arrival time
- FlyTime: flying time
- DepSt: departure station
- ArrSt: arrival station

We will explain Table 2. The first column contains the real flight numbers and the second column contains the renumbered flight numbers. Synchronisation constraints of aircrafts can be found in column 3: Air(i) is the same aircraft, thus as an example, the aircraft of the flight with flight number 1004 is the same as the aircraft of the flight with flight number 1553, 1554, 1333, 1334 and 1873. The same can be done for cockpit (column 4) and cabin (column 5). Departure times respectively arrival times of corresponding flights can be found in column 6 respectively column 7. For the model the actual flying time of a flight is of interest. These flying times can be found in column 8. Departure station respectively arrival station can be found in column 9 respectively column 10, where Out stands for an outstation (i.e. not Schiphol Airport) and * stands for Schiphol Airport.
<table>
<thead>
<tr>
<th>Fln</th>
<th>Ind</th>
<th>ResAir</th>
<th>ResCock</th>
<th>ResCa</th>
<th>DepT</th>
<th>ArT</th>
<th>FlyTime</th>
<th>DepSt</th>
<th>ArrSt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1004</td>
<td>1</td>
<td>Air1</td>
<td>Cock1</td>
<td>Cab1</td>
<td>5:00</td>
<td>7:10</td>
<td>130</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1553</td>
<td>2</td>
<td>Air1</td>
<td>Cock2</td>
<td>Cab2</td>
<td>8:00</td>
<td>9:50</td>
<td>110</td>
<td>*</td>
<td>Out</td>
</tr>
<tr>
<td>1554</td>
<td>3</td>
<td>Air1</td>
<td>Cock2</td>
<td>Cab2</td>
<td>10:40</td>
<td>12:40</td>
<td>120</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1333</td>
<td>4</td>
<td>Air1</td>
<td>Cock3</td>
<td>Cab3</td>
<td>14:20</td>
<td>15:45</td>
<td>85</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1873</td>
<td>6</td>
<td>Air1</td>
<td>Cock4</td>
<td>Cab4</td>
<td>16:35</td>
<td>18:00</td>
<td>85</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1782</td>
<td>7</td>
<td>Air2</td>
<td>Cock5</td>
<td>Cab5</td>
<td>12:00</td>
<td>13:05</td>
<td>65</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1501</td>
<td>8</td>
<td>Air2</td>
<td>Cock2</td>
<td>Cab2</td>
<td>14:15</td>
<td>16:25</td>
<td>130</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1760</td>
<td>9</td>
<td>Air3</td>
<td>Cock6</td>
<td>Cab6</td>
<td>16:20</td>
<td>17:10</td>
<td>85</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1227</td>
<td>10</td>
<td>Air3</td>
<td>Cock1</td>
<td>Cab7</td>
<td>18:40</td>
<td>20:05</td>
<td>110</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1228</td>
<td>11</td>
<td>Air3</td>
<td>Cock7</td>
<td>Cab8</td>
<td>12:15</td>
<td>14:45</td>
<td>70</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1071</td>
<td>12</td>
<td>Air3</td>
<td>Cock7</td>
<td>Cab9</td>
<td>14:05</td>
<td>16:35</td>
<td>85</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1072</td>
<td>13</td>
<td>Air3</td>
<td>Cock7</td>
<td>Cab9</td>
<td>15:15</td>
<td>17:10</td>
<td>65</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1407</td>
<td>14</td>
<td>Air3</td>
<td>Cock7</td>
<td>Cab9</td>
<td>17:20</td>
<td>19:45</td>
<td>85</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1711</td>
<td>15</td>
<td>Air4</td>
<td>Cock1</td>
<td>Cab10</td>
<td>14:25</td>
<td>16:35</td>
<td>125</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1712</td>
<td>16</td>
<td>Air4</td>
<td>Cock1</td>
<td>Cab10</td>
<td>15:15</td>
<td>17:10</td>
<td>125</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1033</td>
<td>17</td>
<td>Air5</td>
<td>Cock8</td>
<td>Cab7</td>
<td>15:00</td>
<td>17:10</td>
<td>90</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1034</td>
<td>18</td>
<td>Air5</td>
<td>Cock8</td>
<td>Cab7</td>
<td>17:25</td>
<td>19:30</td>
<td>95</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1635</td>
<td>19</td>
<td>Air5</td>
<td>Cock3</td>
<td>Cab11</td>
<td>18:55</td>
<td>20:40</td>
<td>105</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1909</td>
<td>20</td>
<td>Air6</td>
<td>Cock9</td>
<td>Cab1</td>
<td>8:30</td>
<td>9:45</td>
<td>75</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1910</td>
<td>21</td>
<td>Air6</td>
<td>Cock9</td>
<td>Cab1</td>
<td>10:35</td>
<td>11:45</td>
<td>70</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1367</td>
<td>22</td>
<td>Air6</td>
<td>Cock10</td>
<td>Cab12</td>
<td>12:40</td>
<td>14:10</td>
<td>210</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1368</td>
<td>23</td>
<td>Air6</td>
<td>Cock10</td>
<td>Cab12</td>
<td>17:00</td>
<td>19:40</td>
<td>240</td>
<td>Out</td>
<td>*</td>
</tr>
<tr>
<td>1018</td>
<td>24</td>
<td>Air7</td>
<td>Cock11</td>
<td>Cab2</td>
<td>4:40</td>
<td>6:30</td>
<td>110</td>
<td>Out</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 2: All necessary information for the example of section 3
### Appendix B

The influence of delays on departure times in air traffic is shown in the table below.

- **$A^{ai}$**

<table>
<thead>
<tr>
<th>180</th>
<th>145</th>
<th>170</th>
<th>120</th>
<th>135</th>
<th>115</th>
<th>160</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
</tbody>
</table>

- **$A^{co}$**

<table>
<thead>
<tr>
<th>180</th>
<th>145</th>
<th>170</th>
<th>120</th>
<th>135</th>
<th>115</th>
<th>160</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
</tbody>
</table>

- **$A^{ca}$**

<table>
<thead>
<tr>
<th>180</th>
<th>145</th>
<th>170</th>
<th>120</th>
<th>135</th>
<th>115</th>
<th>160</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
</tbody>
</table>
Note that in this latter matrix $M$, $\varepsilon$ stands for $+\infty$, because if two flights, which we will call $i$ and $j$, are not connected, the maximum delay of flight with flight number $i$ that not quite reaches flight with flight number $j$ is $+\infty$ and vice versa.
<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
<th>Column 4</th>
<th>Column 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td>Value 5</td>
</tr>
<tr>
<td>Value 6</td>
<td>Value 7</td>
<td>Value 8</td>
<td>Value 9</td>
<td>Value 10</td>
</tr>
<tr>
<td>Value 11</td>
<td>Value 12</td>
<td>Value 13</td>
<td>Value 14</td>
<td>Value 15</td>
</tr>
<tr>
<td>Value 16</td>
<td>Value 17</td>
<td>Value 18</td>
<td>Value 19</td>
<td>Value 20</td>
</tr>
<tr>
<td>Value 21</td>
<td>Value 22</td>
<td>Value 23</td>
<td>Value 24</td>
<td>Value 25</td>
</tr>
<tr>
<td>Value 26</td>
<td>Value 27</td>
<td>Value 28</td>
<td>Value 29</td>
<td>Value 30</td>
</tr>
<tr>
<td>Value 31</td>
<td>Value 32</td>
<td>Value 33</td>
<td>Value 34</td>
<td>Value 35</td>
</tr>
<tr>
<td>Value 36</td>
<td>Value 37</td>
<td>Value 38</td>
<td>Value 39</td>
<td>Value 40</td>
</tr>
</tbody>
</table>

Note: The table continues with more columns and rows.
Designing a Multimodal Coding Scheme for PITA

TRAIL Research School, Delft, December 1999

Author
Ir. R.J. van Vark
Faculty of Information Technology and Systems, Delft University of Technology

Thesis supervisor
Prof. dr. H. Koppelaar
Faculty of Information Technology and Systems, Delft University of Technology
## Contents

### Summary

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Personal Intelligent Travel Assistant</td>
<td>3</td>
</tr>
<tr>
<td>2.1</td>
<td>Seamless Multimodal Mobility</td>
<td>3</td>
</tr>
<tr>
<td>2.2</td>
<td>PITA</td>
<td>3</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Alparon dialogue manager</td>
<td>5</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Dialogue modalities</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Related Work</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>PITA Coding Scheme</td>
<td>10</td>
</tr>
<tr>
<td>4.1</td>
<td>Outline</td>
<td>10</td>
</tr>
<tr>
<td>4.2</td>
<td>Dialogue Phase</td>
<td>10</td>
</tr>
<tr>
<td>4.3</td>
<td>Dialogue Modality</td>
<td>12</td>
</tr>
<tr>
<td>4.4</td>
<td>Dialogue Acts</td>
<td>12</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Topic negotiation</td>
<td>12</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Questions</td>
<td>12</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Contributing information</td>
<td>13</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Feedback</td>
<td>13</td>
</tr>
<tr>
<td>4.4.5</td>
<td>Commitment</td>
<td>14</td>
</tr>
<tr>
<td>4.4.6</td>
<td>Social obligations</td>
<td>14</td>
</tr>
<tr>
<td>4.5</td>
<td>Informational Content</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>Conclusions</td>
<td>16</td>
</tr>
</tbody>
</table>

### References | 17
Summary

Developing dialogue management techniques for next-generation automated information services is the main goal of the project as travel assistants put high demands on dialogue management. They target a large customer group through multiple modalities demanding different dialogue styles based on individual experience and task complexity. Essential in obtaining this goal are a coding scheme for multimodal dialogues and metrics for objectively measuring dialogue aspects. Using these tools, a flexible and adaptive dialogue manager can be developed.

Current dialogue coding schemes are focused on speech-driven systems and have to be extended to cover multimodal dialogues applicable as formalism for PITA dialogue management. This paper describes the design of such a multimodal scheme. Dialogue utterances are coded as comma-separated dialogue acts that code the type of contribution made by the utterance, e.g. a request, feedback, etc. Dialogue acts are performed in a certain dialogue phase and by using one of several available modalities. The latter is important, as some modalities are more error-prone than others are. The informational content of an utterance is coded using a hierarchical tree of information codes.
Developing effective database management techniques is the main goal of the project. The system needs to efficiently store and retrieve large amounts of data. This requires careful consideration of data structures and indexing strategies to ensure high performance and low maintenance costs. Efficient query processing is also crucial for the system's success.

Establishing a robust and scalable database management system can be challenging. It requires a deep understanding of the database's requirements and the ability to anticipate future needs. The system must be adaptable to changes in data volume and complexity. To achieve this, a well-thought-out design and implementation strategy is necessary.

One aspect of the project is database normalization. This involves restructuring the database schema to reduce data redundancy and improve data integrity. This process is critical in ensuring the database's efficiency and reliability.

In conclusion, developing a robust database management system for large-scale data is a complex endeavor. It requires expertise in various areas, including data modeling, query optimization, and system scalability. The project will push the limits of current database management technologies and contribute to the advancement of the field.
1 Introduction

The research programme "Seamless Multimodal Mobility" (SMM) of the Netherlands Research School for Transport, Infrastructure and Logistics (TRAIL) at the Delft University of Technology aims at developing new public transport services for the next century. The new transport services will be developed based on the concept of chain mobility. In this concept, a chain conductor is responsible for the perfect connection between all different transport modalities. The traveller will be transported from door to door without spending time on finding the travel schedule best suited to his needs or worrying about possible delays or calamities while being in transit. Applying the results of the SMM programme to current public transport systems will result in drastic changes in these transport services as new transport modalities and transfer nodes will be introduced and transport services will be provided on a more demand-driven and less fixed basis.

A key role will be played by the Personal Intelligent Travel Assistant (PITA). The PITA is a handheld device that provides ubiquitous communication between travellers and service providers, like the Dutch railways and bus companies. This communication consists of travel planning and reserving capacity before travelling. While travelling, the PITA will guide the traveller by signalling upcoming transfer points and providing information on transfer routes in the nodes. En route, the PITA will also provide the traveller with information concerning delays, calamities and alternative routes, if applicable.

As task complexity, user preferences and traveller surroundings will diversify drastically, multiple modalities will have to be provided for accessing the available information. For example, the complex task of travel planning might require the freedom of a natural language interface, while providing delay information in transit might call for simple but effective text interfaces.

The PITA project aims at developing a prototype service using several modalities: a spoken natural language interface and textual interfaces. Due to the costs of exploiting human operated services, this modality will probably not be taken into consideration for the first versions of the PITA. The prototype service will be based on one dialogue manager that can be interfaced using any of the implemented modalities.

Developing a prototype service requires:

- a multimodal coding scheme as a standard for describing the information dialogues,
• designing and implementing a multimodal dialogue manager to manage information transactions using the multimodal coding scheme,
• examining user requirements and expectations concerning such information services and automated versions thereof.

This article is aimed at designing an effective coding scheme that can be used to represent the dialogues conducted in all modalities. This coding scheme will also be used as a dialogue representation in the PITA dialogue manager.

Background information is provided first in this paper, both on the Seamless Multimodal Mobility programme in general and the PITA project in particular. Related work on coding schemes is discussed next. In the fourth section, the multimodal coding scheme will be presented. Finally, some conclusions will be given and upcoming future work will be described.
2 Personal Intelligent Travel Assistant

This chapter provides background information on the research programme “Seamless Multimodal Mobility” and on the project “Personal Intelligent Travel Assistants” in particular.

2.1 Seamless Multimodal Mobility

Seamless Multimodal Mobility is one of the Delft Interfaculty Research Centres (DIOCs). The research programme was initiated by the Netherlands Research School for Transport, Infrastructure and Logistics (TRAIL).

The research programme Seamless Multimodal Mobility (SMM) aims at the development of building blocks for future seamless multimodal mobility in personal transport. Chain mobility is a solution to prevent the envisioned nightmare of endless congestion and chaotic public transport in the near future. In the concept of chain mobility, a chain conductor is responsible for the optimal connection between trains, buses, taxis and private transport, like cars. The traveller will be transported from door to door without having to worry about transfer times, delays and difficult connections [Bovy et al., 1999].

Seamless multimodal trip chaining will be made possible by sophisticated information and communication technologies as well as real-time transport process control. This will result in effective and efficient multimodal transport services exhibiting an unprecedented high level of quality, which are able to compete with the car.

Multimodal trip chains are achieved by employing a broad gamma of individual (walk, bicycle, car) and public transport modalities (bus, train, ferry, aeroplane). These modalities are organised as a combination of fixed and flexible demand-responsive services. They will adequately cope with the partly unpredictable time-varying access restrictions, quality of transport supply and economy of fares.

The future multimodal transport system will be an integrated mixed, flexible multi-layered network of various types and forms of transportation services that are linked together in intermodal transfer nodes. This service network is supported by physical infrastructure links and nodes and extensive information and communication networks for travellers, service providers, vehicle drivers and public transport operators. Systems like the Personal Intelligent Travel Assistant (PITA) will play a key role.

2.2 PITA

The Personal Intelligent Travel Assistant (PITA) is a handheld device with multimedia capabilities that provides ubiquitous communication between travellers and service providers at any time before or during the trip. Taking into account the recent advances in mobile telecommunication technology, the PITA will be considered to be a sophisticated mobile
telephone providing not only speech interfaces, but also textual interfaces and even wireless Internet access.

In a seamless multimodal transport environment, transport services are no longer based on fixed time schedules only. As a consequence, travellers can never fall back on these fixed time schedules nor can they fall back on fixed transfer stations. Therefore, it is extremely important to provide travellers with adequate and up-to-date information. Using this information, travellers can decide on the travel plan best suited to their needs and they can be promptly informed of changes in their travel plan, like delays. The PITA will integrate several information services, ranging from the complex task of planning a travel schedule for a future trip using different transport modalities to informing the user of transport delays and alternative travel schedules. Depending on the traveller's current surroundings and task complexity, any of the above mentioned dialogue modalities can be used to access the required information. Providing such detailed and personal information to travellers will change public transport systems from impersonal systems aimed at large customer groups into personal systems where the customer gets the impression that the services are particularly designed for him or her personally.

The focus of the PITA project is on developing a multimodal dialogue manager. Dialogue management has become essential in automated information services, both for current speech-driven as well as future multimodal systems. However, current dialogue systems are either fixed and formal, or targeted at a small user group experienced in the application domain. Also, nearly all information services are currently unimodal.

The PITA project invests in dialogue management since:

- travel assistants are targeted at a large group of customers demanding different dialogue styles based on the individual preference and experience of users,
- dialogues have to be adaptive depending on the complexity of the task and success of the dialogue so far, and
- the Personal Intelligent Travel Assistant is intended to be multimodal.

The first versions of the PITA are not supposed to be *simultimodal*¹, i.e. the different modalities of the travel assistant are not supposed to be used simultaneously. The same information service can however be accessed by using different modalities sequentially. For example, the user can plan his trip by using speech, while information updates can be sent using text interfaces. Also, certain modalities might be favourable in certain surroundings. For instance, text interfaces are not suited to be used in a car environment.

¹ Simultimodal is a contraction of simultaneous and multimodal.
2.2.1 Alparon dialogue manager

The dialogue manager will be based on the dialogue manager of the Alparon system. Alparon is a research group at the Delft University of Technology working on dialogue management for information and transaction services. The Alparon dialogue manager has been especially designed for spoken language input [Van Vark et al., 1997a], but the design process has always taken into account a possible inclusion of other modalities in the future. As a result, most of the modules making up the Alparon dialogue manager are modality-independent. Only the disambiguation module, the first module which takes care of disambiguating user input, and the dialogue act generator, the final module which takes care of translating the system response into suitable dialogue acts, are dependent on the modality applied in the information service. However, even their dependence on dialogue modality is limited as the Alparon dialogue manager uses dialogue acts to communicate with other natural language components in the Alparon system.

Other modules in the Alparon dialogue manager take care of context updating, dialogue updating and response generation [Andeweg, 1999]. Besides these modules, the dialogue manager also consists of a number of blackboards, which store all information relevant to the ongoing dialogue. Amongst others, there are blackboards for dialogue context, dialogue history, control information, and information to be presented to the user.

Using the Alparon dialogue manager for the PITA information service calls for several modifications to the dialogue manager. First, the dialogue manager has to be made fully multimodal. This does not only apply to making the current coding scheme suitable for other modalities, but also to enhancing the coding scheme with dialogue phenomena which might not be found in spoken language or in the original Alparon domain. Secondly, adaptive dialogue control strategies are needed to suit the wide range in task complexity, user experience and user environment.

2.2.2 Dialogue modalities

As mentioned above, several dialogue modalities are foreseen in the Personal Intelligent Travel Assistant. These modalities can be described using several characteristics, for instance complexity of the interaction and environment in which the modality is best applied. For the moment, the following dialogue modalities are foreseen in the Personal Intelligent Travel Assistant.

Short Message Service (SMS)

Short Message Service is an information service that can be found on many platforms for mobile telephony. It can be used to sent text messages

---

2 In fact, the dialogue manager for the PITA system will be the next generation Alparon dialogue manager. PITA will be the information service for which this dialogue manager will be developed.
with a maximum size of 160 characters to other mobile subscribers or information servers. Human-human conversation based on SMS usually consists of free formed text, like "Hi, let's meet at the Chinese restaurant at 8 p.m. Robert".

Accessing the information services is done by sending complex command-like text messages to the information server. Usually, no on-line help is available on the command syntax and subscribers have to memorise all possible commands and abbreviations. For instance, to query a commercially operated information service in the Netherlands for an update when the stock price of KPN rises above 44.00, the command would be: "TRIGGER AE KPN+44.00". SMS is especially popular among young people as it provides an easy and cheap platform to schedule appointments wherever you are.

Although sending SMS messages back and forth can constitute a dialogue, SMS is best suited to straightforward initiation-response scenarios including short digressions when unclear or ambiguous information is found. In the PITA environment, SMS is best suited for providing information on upcoming transfers and delays in transit. SMS dialogues consisting of multiple turns are extremely time consuming.

Due to the text interface, SMS is not suited to be used when driving a car or riding a bike/motor. However, SMS is suited to be used in transit, especially in crowded environments, where social constraints might limit the use of spoken interfaces.

**Wireless Application Protocol (WAP)**

The Wireless Application Protocol (WAP) combines the best of two rapidly evolving network technologies: mobile data and the Internet. Most of the technology developed for the Internet has been designed for desktop and larger computers, medium to high bandwidth and reliable data networks. WAP is a protocol set making Internet-like information services available for handheld devices.

Compared to SMS, information services using the Wireless Application Protocol can be much more complex while the interface is still straightforward to use for the inexperienced user. The WAP dialogue is, however, very constraint and system-driven. The user has to answer the questions asked by the system in order to gain access to the information he or she wants. Diversions of the system scenario are hardly possible.

The environmental constraints on WAP are approximately the same as for SMS; using a WAP interface while riding a bike or driving a car is virtually impossible. A WAP interface seems to provide a user-friendlier interface than SMS driven interfaces, even suitable for more complex information services. However, the Wireless Application Protocol is currently being developed and the first WAP-driven information services are just being introduced. Also, WAP-enabled mobile telephones are expected to be first introduced in the Netherlands late 1999.
Automated spoken language

Automated spoken language interfaces are used to interface information systems by using the human voice. Spoken language interface can be used in virtually any environment, although adverse environments, like driving a car at high speed, cause a significant drop in performance. Over the last decade, speech recognition has evolved drastically making it applicable to a wide range of applications, among which (complex) information retrieval tasks.

Despite the recent advances, automated spoken language processing still has several drawbacks. In contrast to the suggested freedom of speech, speech recognisers apply suitably small vocabularies to prevent large numbers of recognition errors, certainly when applying speech recognition in a mobile telephony environment. Also, extensive confirmation scenarios are still needed due to the occurring recognition errors and misinterpretations due to out-of-vocabulary errors. Adding the social limitations of using speech technology in public, limits the potential use of automated speech processing in PITA.

Spoken language interfaces are especially suited to be used in more complex tasks, which cannot be interfaced by using more straightforward text interfaces, because the amount of information would be overwhelming when using a text interface. In the Personal Intelligent Travel Assistant, automated speech processing can be applied for travel planning as it is the most complex task. Also, travel planning can be performed in a more isolated environment before taking part in the actual transportation process.
3 Related Work

This chapter discusses related work on dialogue coding schemes. Most coding schemes have originally been developed to analyse human-human or human-computer dialogues, while only a few schemes have especially been developed to perform dialogue management.

The basic assumption of a plan-based theory for dialogue modelling [Cohen, 1994] is that the linguistic behaviour of agents in information dialogues is goal-directed. To reach a particular state agents use a plan, which is often a small variation of a standard scenario. The structure of a plan resembles the dialogue structure. Dialogue acts in PITA dialogues are also assumed to be part of a plan and the listener is assumed to respond appropriately to this plan and not only to isolated utterances.

Carletta et al. developed a coding scheme for task-oriented dialogues existing of three levels [Carletta et al., 1995]. At the highest level dialogues are divided in transactions or subdialogues corresponding to major steps in the participants’ plan for completing the task. At the intermediate level conversational games consist of initiations and responses. A conversational game is in fact a sequence of utterances, starting with an initiation and ending with the requested information being transferred or acknowledged. At the lowest level, conversational moves are utterances that are initiations or responses named according to their purpose.

In the VERBMOBIL project, researchers developed a taxonomy of dialogue acts [Alexandersson and Reithinger, 1995]. To model the task-oriented dialogues in a large corpus they assumed that such dialogues can be modelled by means of a limited but open set of dialogue acts. They defined 17 dialogue acts and semi-formal rules to assign them to utterances. These assignment rules served as a starting point for the automatic determination of dialogue acts within the semantic evaluation system. From the analysis of an annotated corpus of 200 dialogues they derived a standard model of admissible dialogue act sequences.

While in the first phase of the VERBMOBIL project the scenario was restricted to appointment scheduling only, it has been extended to include travel planning in the VERBMOBIL-2 project. The VERBMOBIL-2 coding scheme includes a new set of dialogue acts and dialogue phases [Alexandersson et al., 1998]. The new set of dialogue acts is organised as a decision tree containing 29 dialogue acts for coding utterances. Researchers in the VERBMOBIL-2 project distinguish between controlling the dialogue, task management, and information transferral. The dialogue acts are used to express the primary communicative intention behind an utterance. As a consequence, utterances are annotated with as few dialogue acts as possible, but the annotation is not necessarily restricted to one dialogue act for every utterance.

Another approach of modelling dialogues [Bunt, 1995; Rats, 1996] is based on the observation that in goal-directed dialogues, the information is exchanged in such a way that the dialogue contains one or more topical
chains. These are sequences of utterances that all communicate information about the same topic (topic packets). Successive topics are usually related to each other. Topic transitions in a dialogue are modelled as movements across these topic packets. The local relationship between topics can be used in knowledge based systems in a dialogue manager. Since the introduction of plan theory, researchers criticised a plan-based approach. For example, Dahlbäck noticed that in advisory dialogues it is possible to infer the non-linguistic intentions behind a specific utterance from knowledge of the general task [Dahlbäck, 1995]. In an information retrieval task, this inference can be very difficult, however, in order to provide helpful answers this information is not needed frequently. Since the PITA dialogues are strongly goal-directed, intentions are not very important. In a demonstrator for the German railways and a recently developed service for the Dutch railways, intentions are deliberately neglected. This system is strongly goal-directed as well [Oerder and Aust, 1994].

The Alparon coding scheme was used at the Delft University of Technology to analyse information transfer strategies in public transport information dialogues [Van Vark et al., 1996b; Van Vark et al., 1997b]. The coding scheme consists of a set of dialogue moves used to code the intention behind an utterance and a highly hierarchical and detailed coding scheme to annotate the information transferred in the utterances. The sequence of dialogue phases followed during a dialogue was used to develop a taxonomy of information retrieval dialogues [Van Vark et al., 1996a]. The PITA coding scheme is a refinement of the Alparon coding scheme.

Since 1996, researchers organised in the Discourse Resource Initiative (DRI) [Traum and Nakatani, 1999] have been working on a standardisation effort for annotating communicative acts in dialogue. Their DAMSL, Dialog Act Markup in Several Layers [Core and Allen, 1997], consists of three layers: forward communicative functions, backward communicative functions, and utterance features. Each layer allows multiple communicative functions of an utterance to be labelled. The forward communicative functions consists of a taxonomy in a similar style as the actions of traditional speech act theory [Searle, 1969; Searle et al., 1980; Traum, 1999]. The backward communicative functions indicate how the current utterance relates to the previous dialogue, such as accepting a proposal, confirming understanding, or answering a question. The utterance features include information about an utterance's form and content, such as whether an utterance concerns the communication process itself or deals with the subject at hand. The DRI standardisation effort shows promise but inter-annotator reliability has to be improved further. The main cause for this lack in reliability is the high number of dimensions on which utterances can be annotated using DAMSL.
4 PITA Coding Scheme

The goal of the PITA research project is to develop prototype services for providing online traveller guidance using multiple modalities. The main point of interest is how information is transferred between dialogue participants using the implemented modalities. This knowledge will be applied in the development of a multimodal dialogue manager. As it also focuses on information transfer, the design of the PITA coding scheme is based on the Alparon coding scheme [Rothkrantz et al., 1996; Van Vark et al., 1997b] and the experience obtained by applying that coding scheme to spoken language dialogues in the public transport domain. Results of the Discourse Resource Initiative's standardisation effort [Core and Allen, 1997; Allen and Core, 1997] and the VERBMOBIL-2 project [Alexandersson et al., 1998] have also been used to improve the coding scheme.

4.1 Outline

Dialogues consist of turns that normally alternate between agents. These turns consist of one or more utterances that correspond to grammatical sentences or text messages. These utterances are coded as dialogue acts listed in our coding as comma-separated terms.

A dialogue act takes place in a certain phase of the conversation that indicates whether the system or the client has the initiative. This phase is coded as an argument of the dialogue act, because dialogue acts play a more important part in managing information retrieval dialogues than phases. Therefore we prefer to tag dialogue acts with the phases in which they take place, instead of placing dialogue acts hierarchically lower than phases.

As the PITA project involves multiple modalities, these have to be part of the coding scheme also. They are coded similarly to phases, i.e. once for every dialogue act.

Other arguments of dialogue acts are symbols used to code information, like Location, Time, etc. Information codes can have layers of subcodes containing more detailed information.

4.2 Dialogue Phase

Dialogues can be divided into several phases or subdialogues that accomplish one step in retrieving the desired information. Phases are typically a number of consecutive topic packets [Rats, 1996] used to complete such a step in the dialogue. Examples of subdialogues modelled by phases are defining which information is needed or presenting the information to the client.

In our coding scheme, six phases can be distinguished. During the Greeting phase at the beginning of the dialogue, both parties exchange greetings and might introduce themselves.
The **Opening** phase is used to introduce the topic of conversation, i.e. it is used to "open" the information exchange. In many dialogues, the topic is introduced implicitly, e.g. "Your train has a delay of 7 minutes. Your connecting train will wait in Rotterdam instead of Delft."

In the **Information** phase the traveller has the initiative and tries to describe his information need. All parameters that contribute to this need for information can be altered during this phase. In many dialogues there is no explicit question (e.g. 'Tomorrow I have to be in Rotterdam'). When the information given by the customer is insufficient to find a solution, the system can elicit further information by posing supplementary questions.

During the **Presentation** phase the system presents the information asked for by the traveller. This information does not have to be requested in the current dialogue. For instance, the system might update the traveller on a delay for a travel scheme he defined in a previous dialogue. In this phase, the system has the initiative and often presents the information in small pieces to make it easier for the client to memorise or to write down the information. The customer usually acknowledges each piece of information.

The **Closing** phase is used to "close" the information exchange on the current topic of conversation. This can be done either implicitly or explicitly. When done implicitly it usually coincides with the goodbye phase. When done explicitly, it is usually followed by opening another topic of conversation, i.e. the caller wants to satisfy another information need.

The last phase is the **Goodbye** phase, which is used to end the conversation by exchanging thanks and saying goodbye.

Looking at these phases, task-related information is exchanged in only four of them: Opening, Information, Presentation, and Closing. The other phases serve social goals, because information dialogues have to observe norms and conventions of both agents' cultures [Bunt, 1995], although not equally strong for all modalities.

Phases are important aspects of our coding. A dialogue phase is a good indication of the flow of information at a certain time in the dialogue and it denotes which agent has the initiative. For instance, during the Greeting phase both agents present themselves and in some case they also give information about their residence, while in the Query phase the client has the initiative and presents most information.

Phases are not annotated automatically during the automatic recognition process. Dialogue phases are derived by the dialogue manager and used to decide on the best dialogue strategy. Dialogue phase is also a useful aid in analysing information dialogues and it is especially suited as a measure of dialogue complexity [Van Vark et al., 1996b].
4.3 Dialogue Modality

Dialogue modality is important as it influences the dialogue strategy drastically. For instance, it is more time-consuming to provide information using WAP or SMS when compared to spoken language, but automatic recognition yields better results and less strict verification strategies can be applied than in the case of spoken language. Also, information density is usually much higher in textual interfaces. For example, in a WAP dialogue the system can ask for more parameters in one utterance than in a spoken language dialogue.

The modality used to interact is coded as an argument to the dialogue act. In this way, the same coding scheme can be used for sequential multimodal dialogues, i.e. dialogues in which multiple modalities are used but not simultaneously. The codes for dialogue modality simply represent the modalities described in the previous section, i.e. WAP for Wireless Application Protocol, SMS for Short Message Service, and ASP for Automated Speech Processing.

4.4 Dialogue Acts

Several types of dialogue acts can be distinguished in the coding scheme. These types are used for different purposes: topic negotiation, questions, contributing information, feedback, commitments, and necessary social behaviour. These groups will be discussed below.

Unlike in DAMSL, the coding scheme proposed by the Discourse Resource Initiative, utterances are annotated with as few dialogue acts as possible. However, when multiple dialogue acts apply to an utterance and no dialogue act covers the function of the combination of these dialogue acts, the utterance is annotated with multiple dialogue acts. This is similar to the coding applied in VERBMOBIL [Alexandersson et al., 1998].

4.4.1 Topic negotiation

The dialogue acts in this group cover explicit topic handling. Topic handling was not part of the Alparon coding scheme. Topics are explicitly introduced by using the dialogue act Init. The described utterance contains an information element describing the topic to be discussed. A Close dialogue act ends the information exchange concerning a certain topic. In contrast to an Init utterance, the Close dialogue act does not contain an explicit reference to the handled topic. However, the utterance ends the topic handling explicitly. Usually, explicit topic closing is acknowledged by the dialogue partner.

4.4.2 Questions

In the previous Alparon coding scheme, there were three question-like dialogue acts: question, which asked for information, reconfirmation, which asked for a repetition of information, and check, which checked information the speaker had reason to believe but which had not been mentioned explicitly in the dialogue. In the PITA coding scheme, four
different request types also used in the VERBMOBIL project are used, supplemented by one additional question. 

Request_suggest is the basic way of asking for information. Request_suggest obligates the dialogue partner to respond further in the dialogue by supplying the requested information if it is available. Request_choice also obligates the dialogue partner to respond to the request. However, this dialogue act provides several choices from which the dialogue partner should choose, instead of an open-ended request. The Request_choice is often found in WAP dialogues where the interface is especially suited to provide an extensive list of choices to choose from. Request_clarify asks for a clarification of a piece of information. This information must have already been mentioned in the dialogue. Request_clarify also includes requests for repetitions of information provided earlier in the dialogue.

Request_comment dialogue acts are basically yes/no-questions asking the dialogue partner whether he/she agrees with the mentioned information. An explicit topic closing is often combined with a Request_comment act. The final question-like dialogue act is Request_commit. An utterance coded as Request_commit asks the partner to commit him/herself to some future action. In the PITA domain, this dialogue act will be mainly used to commit the customer to a capacity reservation like a taxi reservation.

4.4.3 Contributing information

Three ways of providing information to the dialogue context are foreseen in the coding scheme. Suggest introduces information that has not been mentioned in the dialogue so far. As the name indicates, the information is a suggestion and can be rejected by the dialogue partner. In most cases, the dialogue partner responds to a Suggest by providing positive or negative feedback. In contrast to Suggest, Inform provides factual information. The dialogue partner can however comment on this information. This dialogue act is mainly used for providing factual information on arrival or departure times, travel delays, etc. The last way of providing information is Clarify. This is either a response to a Request_clarify and tries to improve the partners understanding of a certain piece of information, or it contains a repetition of information that has already been mentioned in the dialogue.

4.4.4 Feedback

As a response to a Request_comment or a Suggest/Inform act, dialogue participants supply feedback. This feedback can be either positive or negative. In human-human dialogues other values can be provided indicating a finer grade of (dis-)approval. These values are, however, seldom found in human-computer interaction.
In the Alparon coding scheme, feedback was coded using one dialogue act only (acknowledgement) including information codes indicating positive agreement, partial agreement or disagreement. In the current scheme, feedback will be coded using four dialogue acts.

Accept is the dialogue act used to show agreement on previously mentioned information or actions. Accept is only used to code utterances indicating total agreement like "yes", "sure", "that’s ok", etc. When (part of) the information is referenced in the utterance, it is coded as an Accept_partial. In this case, the dialogue act contains one or more information elements indicating the information the dialogue partner agrees upon. Even when all information of the previous utterance is mentioned in the agreement utterance, it is still coded as an Accept_partial.

Negative feedback is coded as Reject. Again, this dialogue act is only used to deny all information from the previous utterance without referring to the information itself. When referring to the information and thus rejecting parts of it, the dialogue act Reject_partial is used.

4.4.5 Commitment

Commitments are included in the coding scheme to allow capacity reservation and online traveller guidance. In order to facilitate both services, the system needs to be absolutely sure that the customer is going to make the negotiated travel plan, i.e. the customer has to be committed to the travel plan. This type of dialogue act was not present in the Alparon coding scheme, as the Alparon domain only handled information retrieval dialogues. Whether customers followed up on the provided information was not relevant.

Offer is used for offering to perform a certain action, i.e. taking a specific train or being at a specified location at a specified time. An Offer has to be accepted by the other dialogue participant to commit him to the specified action. Offer contains one or more information elements indicating the action being offered.

Using a Commit, a dialogue participant explicitly commits him/herself to perform the specified action. A commit is usually a response to a Request_commit act and the action to be performed can be referenced anaphorically.

4.4.6 Social obligations

The coding scheme contains two dialogue acts that are motivated by social behaviour only. This explicit social behaviour is most often found in spoken language dialogues.

When a Greet act is found, it is the first utterance in a dialogue. It is used to start the conversation. Optionally, the dialogue participant introduces himself by mentioning his name.

Bye is the final dialogue act of a conversation. It shows the dialogue participant’s willingness to end the dialogue. Bye can be performed
implicitly by a hang-up of the dialogue participant. When Bye is performed explicitly it is usually accompanied by showing gratitude for having used the information service.

4.5 Informational Content

The informational content of the PITA coding scheme is based on the analysis of the public transport information corpus at the Delft University of Technology. This analysis resulted in the Alparon coding scheme. This section provides an overview of the information codes provided in this coding scheme; for a detailed description, see [Van Vark et al., 1996a]. For the personal assistant, several small additions have been made to these information codes, mainly to cover the wider domain and topic management.

Information codes are used to code what type of information was transferred in every dialogue act described above. Coding of information was done hierarchically to facilitate automated processing of coded dialogues. An example is Location, followed by an information subcode, e.g. Departure. Arguments to this information code are items like Street or City. In a few cases, arguments have subarguments.

**Topic** concerns information needed to perform explicit topic management. For now, the topic can either be Planning, Guidance, or Delay.

**Location** contains information about locations mentioned in the dialogue. This can be the arrival and departure place, but also intermediate stations or directions. Typical arguments for location are City, Station, etc.

**Time** concerns the general time setting of the query. This includes arrival or departure time and date of travelling, but also relative time settings like the next or last connection. Arguments for this code are typical time and date expressions.

**Route Information** covers the type of journey to be made. This information ranges from Transport Type (e.g. Bus or Train) and Connection Type (e.g. Direct or Nonstop) to Transfers, Travelling time and Delay time.

**Timetable** concerns all information producing insight into the general set-up of the timetable belonging to a specific journey. This can be information of a regular timetable (like 'twice every hour') as well as information concerning changes in regular timetables as an effect of irregularities.
Conclusions

The coding scheme to be used for managing Personal Intelligent Travel Assistant dialogues is presented. This coding scheme is a refinement of the Alparon coding scheme. The refinement is based on the experience gained by applying the Alparon coding scheme to public transport information dialogues and results from efforts of the Discourse Resource Initiative and the VERBMOBIL-2 project. Using the described design, the applicability of the coding scheme to the PITA information service can be assessed. First, suitable dialogues from the Delft University of Technology OVR corpus can be annotated using the PITA coding scheme. Special attention needs to be paid to the information codes used as it is difficult to list all information that will be applied in the PITA project beforehand. Secondly, the Alparon dialogue manager needs to be modified in order to incorporate the new coding scheme. The dialogue manager can then be used to build prototypes for the PITA information services. The first services will be operated using the Wireless Application Protocol and spoken language. Only then can the validity of the new coding scheme be fully assessed.

When using different modalities, it has to be investigated whether all modalities can be directly coded using the PITA coding scheme. Another possibility might be to use several modality coding schemes which can be translated to the core dialogue management coding scheme.

Future work

Once the validity of the coding scheme has been guaranteed, focus can be shifted towards new dialogue management strategies for information dialogues. These new strategies must lead to flexible and adaptive dialogue control in order to facilitate the wide range in task complexity, user experience, and interface functionality in the domain of Personal Intelligent Travel Assistants.
References


• Core, M. and Allen, J. (1997). Coding Dialogs with the DAMSL Annotation Scheme. In AAAI Fall Symposium on Communicative Action in Humans and Machines, Boston, MA.


---

Designing a Multimodal Coding Scheme for PITA
Appendix

The Netherlands TRAIL Research School
Crossroads of theory and practice
There are few concepts in the world that have brought greater benefit to mankind than transport - cars, trains, aeroplanes, ships and trucks carrying people and goods to satisfy a myriad needs across the world. But they have only been able to offer their benefits thanks to the combined science of transport, infrastructure and logistics - the complex activities that permit transport to function and are so essential to efficient productive and consumer societies. Without transport linked to infrastructure and logistics our modern standard of living would be unthinkable.

Head start
The Netherlands is today truly at the transport crossroads of the world. The country is a main entry and exit to Europe, it has a long tradition in foreign trade, talents in infrastructure were learned defending its low-lying lands from water, it has an open national economy and high standards of professional education. Notwithstanding such a head start the Netherlands realises that its prosperity still depends on courageous investment, pre-eminence in transport and the constant upgrading of knowledge. The Netherlands Research School for Transport, Infrastructure and Logistics, TRAIL, serves as a national and increasingly international meeting point, at the crossroads of transport theory and practice.

TRAIL
a research centre and postgraduate school in transport, infrastructure and logistics - in service of society worldwide

Knowledge and ability
When one looks to the future success of national economies, knowledge is the key. It is knowledge across a multitude of domains and the ability to apply this that will forge societal and commercial success. A primary capital asset for the future thus lies in the quality of citizens, in the quality of the knowledge infrastructure, in developing and disseminating new knowledge and applying it. Companies and governments realise this; both are giving high priority to the development of the knowledge infrastructure. TRAIL plays an essential role in the process.

Crossroads
At the crossroads of theory and practice means expanding and intensifying available knowledge; it means greater attention to innovation by encouraging investment in research and development. It means scale enlargement, improved cooperation and integration, and a greater role for telematics. It means linking the demand for and supply of knowledge more effectively than in the past. One of the stimulants in this process is optimal cooperation and interaction between knowledge institutions and the companies that apply knowledge in practice. This is TRAIL's position in the world of transport, infrastructure and logistics: the crossroads of knowledge and practice.
What is TRAIL?
TRAIL is the Dutch academic research institution targeting worldwide developments in transport, infrastructure and logistics, and combining top-level education and research. A knowledge institute in which the Delft University of Technology (DUT), the Erasmus University Rotterdam (EUR) and the University of Groningen actively participate. Faculties of economics, business administration, mathematics and technology are among those combining their strengths to create unprecedented levels of knowledge synergy. The purpose: using experience and talents combined with those of clients to offer solutions to challenges in these fields across the world. TRAIL Research School is a breeding ground for innovative solutions.

TRAIL as a source of knowledge
TRAIL links postgraduate training, research, and the application of this in society; tasks demanding energy in the reanalysis of old work, the reassessment of old beliefs and the ordering and structuring of a vast resource of existing knowledge. It is requiring the creation of entirely new knowledge. It demands linkage of complex subject matter and an attitude of integral engineering. Problems in the field of transport, infrastructure and logistics are generally so complex that they can no longer be solved from one point of view. So the TRAIL Research School uses the multi-disciplinary approach; the integrated competence of many specialists in and beyond the domain transport, infrastructure and logistics. TRAIL strongly emphasises the coupling of science and practice. It represents a key link between knowledge and application, and between academia, business and government. An international network links TRAIL with other international knowledge institutes in the same field.

TRAIL in practice
TRAIL operates on the basis of creating a multidisciplinary team optimally suited to a challenge in view. Client consortia may be created to back larger more expensive projects. Cooperation between clients and consultants is close and regular. TRAIL has worked in such an integrated manner on matters such as the development of mainports Rotterdam and Schiphol, transport in the western Dutch conurbation, vehicle use and guidance of traffic, specifically in congestion-prone areas and the problems surrounding urban freight transport. Many subjects are highly international in character.

A specific example was TRAIL's contribution to the 'IncoMaas MasterPlan'; an original approach to the handling of container transhipment in Rotterdam to further improve the competitive position of the largest port of the world. Proposals were presented for new physical layouts, the integration of intermodal transport methods, fully automated storage and transhipment, improvement of connections to the hinterland and better accessibility to the western conurbation. Many are being implemented. In the area of people transport work is being undertaken on the creation of a flexible, demand-driven collective passenger transport system. Here the electronic highway is likely to play as important a role as the public highway.
Business, government and TRAIL

In TRAIL's approach, the input of government, business and the people on the ground is essential. TRAIL teams do not only listen very carefully, they contribute with the formulation of questions, development of ideas and contribution to implementation. TRAIL also offers direct access to some of the world's most competent transport specialists. The Public-Private Academic programme, in which businesses and institutions participate directly, allows partners to have immediate access to all TRAIL activities. They establish direct contact with leading researchers and research groups working on tomorrow's topics and are assured of access to the international knowledge circuit. Participants regularly receive information from TRAIL on new R&D projects. In short, partners are assured of a direct line of contact to a unique source of knowledge on transport, infrastructure and logistics.

Next steps

If the reader would like to know more about TRAIL, the organisation, its work and specific opportunities for a company or institution, please do not hesitate to call. A member of TRAIL management will be delighted to tell you all you want to know.

Telephone : +31 (0)15 278 6046
Telefax : +31 (0)15 278 4333
E-mail : mailbox@TRAIL.tudelft.nl
Internet : http://TRAIL.tudelft.nl

The face of TRAIL

Ratified by the Boards of both universities, TRAIL officially started on 1 January 1994 as the Netherlands Research School for Transport, Infrastructure and Logistics offering participation to all specialist departments. There are currently more than 230 researchers active at TRAIL, of which 125 qualifying to Ph.D. level. TRAIL is accredited by the Royal Netherlands Academy of Arts and Sciences.

TRAIL clients include the European Union, various Dutch ministries, state and provincial governments, the national Department of Transport, Waterways and Public Works, the Netherlands Institute for Natural Sciences, several companies and institutions (Albert Heijn, Amsterdam Airport Schiphol, Centre for Transport Technology, City and regional public transport companies, DSM, DTO, Dutch Railways, Europe Combined Terminals, Gasunie, KPN, Koninklijke Pakhoed, the Port of Amsterdam and Rotterdam, Railned, Van Ommeren, VBA, Hollandia I.M., Heineken, Van Gend & Loos). Details on reference projects are available.
TRAIL management and Programme Board
Prof.dr.ir. P.H.L. Bovy Drs.ing. A.L. Loos
Scientific Director and ‘Seamless Multimodal Mobility’

Prof.dr.ir. R. Dekker
‘Vehicle Control Technology Impacts’
‘Logistic Process Control’
‘Gloria’

Prof.dr.ir. J.J.M. Evers
‘Planning and Maintenance of Infrastructure’
‘Logistic Process Control’

Prof.dr.ir. R.E.C.M. van der Heijden
‘Chain Management and Control’
‘Seamless Multimodal Mobility’

Prof.dr. M.B.M. de Koster
‘Freight Transport Automation & Multimodality’
‘Transport Policy Analysis’
‘Dynamic Traffic Management’

Prof.dr.ir. J.A.E.E. van Nunen
Prof.dr. G.J. Olsder
Prof.dr.ir. H. Priemus
Prof.dr.ing. G.R. Teisman
Prof.dr. H.J. van Zuylen

Industrial Advisory Council
M.C.J. van Permis
Mr. W. de Boer
Chairman, Siemens Nederland N.V.
Ministry of Transport and Water Management (V&W)
Europe Combined Terminals (ECT)
Ministry of General Affairs (AZ)
Haagse Tramweg Maatschappij (HTM)
Amsterdam Airport Schiphol (AAS)
Port of Rotterdam (GHR)
Hollandse Beton Groep NV (HBG)
The Netherlands Railways (NS)
ConneKt

Mr. W. de Boer
Ir. W. den Dulk
Mr. L.A. Geelhoud
Ir. G.A. Kaper
Drs. J. Krul
Mr. S.W.A. Lak
Ir. B.J.G. van der Pot
Ir. T. Regtuijt
Prof.dr.ir. G.J. Wormmeester

Board of Founding Faculties
The Board is made up of the deans of university faculties.

Prof.dr. H. Bart
Prof.ir. H. Beunderman
Prof.dr.ir. R. Dekker
Prof.mr. H. de Doelder
Prof.dr. P.J. Idenburg
Prof.ir. Th. de Jong
Prof.dr.ir. J. van Katwijk
Prof.ir. Klein Woud
Prof.dr.ir. P.B. Lehning
Prof.ir. H.J. Overbeek
Prof.dr. P.H.A.M. Verhaegen

Prof.dr.ir. P.H.L. Bovy
Drs.ing. A.L. Loos
Chairman (a.i.), Faculty of Economic Sciences
Faculty of Architecture
Faculty of Economic Sciences
Faculty of Law
Faculty of Technology, Policy and Management
Faculty of Aerospace Engineering
Faculty of Information Technology and Systems
Faculty of Design, Engineering and Production
Faculty of Social Sciences
Faculty of Civil Engineering and Geosciences
Faculty of Business Administration

Scientific director TRAIL Research School
Managing director TRAIL Research School

1 Delft University of Technology
2 Erasmus University Rotterdam
3 University of Groningen
TRAIL Conference Proceedings

A series of The Netherlands TRAIL Research School for conference papers on transport, infrastructure and logistics.


Seminar TRAIL Research School, Recent Advances in Traffic Flow Modelling and Control, September 1999, TRAIL Conference Proceedings, P99/2, The Netherlands


Sales and distribution:
Delft University Press
P.O. Box 98
2600 MG Delft
Telephone: +31 (0)15 278 32 54
Telefax: +31 (0)15 278 16 61
The Netherlands TRAIL Research School
Delft University of Technology
Erasmus University Rotterdam
University of Groningen

Kluyverweg 4
P.O. Box 5017
2600 GA Delft
The Netherlands
Telephone: +31 (0)15 278 60 46
Telefax : +31 (0)15 278 43 33
http://www.TRAIL.nl