A network-wide control method for freeways

Reducing the travel times in large traffic networks by efficient coordinated and integrated control

January 2012

Odile De Vito
Preface

With this document, I proudly present the last phase of my MSc at the University of Technology Delft. It has been an interesting, joyful and instructive ride since I arrived in Delft starting my BSc at the faculty of Technology, Policy and Management. I was very lucky to discover my interests in quantitative research on time, which motivated me to switch to the faculty of Civil Engineering for the MSc Transport & Planning. One of the first courses I followed was about Traffic Flow Theory, and immediately I knew for sure that I wanted to graduate on this subject. This brought me to a graduation internship at TNO mobility, where I have researched the possibilities to control large freeway networks online. This research has resulted in a new coordinated and integrated control method for freeway networks, with a low computation complexity.

I could not have delivered this document without the help of supervisors, colleagues, friends and family and therefore I would like to thank many persons. At first Andreas Hegyi, my daily supervisor, who told me when choosing this subject that it would not be easy but that he would help me whenever I needed. And he did! Our meetings gave me new thoughts and solutions (and many more questions) but most important was the motivation and confidence I gained every time: thank you so much Andreas! Next, I want to thank my supervisor at TNO, Freek Faber. He did not only review my documentary critically and gave me a new perspective on the subject but also made me feel at home at TNO. The chairman of my graduation committee, professor Serge Hoogendoorn, was the one that triggered my interest in Dynamic Traffic Management at first. Not only could he challenge me and did he help me out with difficult choices, I really enjoyed our chats! Bart De Schutter of DCSC was added to the group at last, but really helpful for explaining me the basics of traffic control: thank you for supervising me although you were so busy! The fifth member of my committee I want to thank is Paul Wiggenraad, for his feedback and for convincing me to choose this master three years ago.

The last eight months, I have spent most days in Delft at TNO mobility and I want to thank all my colleagues there for their help and company. Especially I want to express my gratitude to Taoufik, Paul, Damir, Martijn, Maaike, Simeon and my intern colleague Jos.

At the university I was adopted by the Trafcon research group, which have given my very useful feedback on my research and arranged many interesting meetings. Special thanks to Ramon, Thomas, Mohammad, Erik-Sander and Michiel Bliemer.

Last but not least I want to thank my family and friends for their support. My parents, for always having faith in me, and my sister Laura for her hospitality. My housemates Daphne en Tienieke for their advice and spirit. Michiel for all the meals he cooked for me and for cheering me up with his sarcasm. Margriet, Karin and Sara for all the lovely evenings of dining and wining. And Arjen for his reflection and ability to make me totally relaxed. I feel lucky to have you all in my life!

Odile De Vito
Delft, January 2012
Summary

The increasing traffic demand exceeds the current road capacity at several locations, especially during peak hour, and results in congestion. Congestion has negative consequences for the economy, the environment and the human experience: the social welfare decreases. This problem can be solved by adding capacity to the network (which is expensive and not always possible), reducing the traffic demand or by better utilization of the current network: applying dynamic traffic management (DTM) measures to the traffic network. The last approach is proven to be successful: DTM measures can reduce the lost vehicle hours, but only when these measures are well controlled.

Several control methods have been designed for the control of DTM measures in both theoretical studies as practical implementation, such as the local control of ramp metering by the ALINEA algorithm and the AMOC algorithm for optimal control of ramp metering at freeway corridors. However, these control methods leave opportunities for improvement. It is proven that a coordinated network-wide approach, where there is not focussed on solving a single bottleneck but on the performance of the whole network, can improve the network performance compared with decentralised control. Unfortunately, it is difficult to find the optimal solution in real-time, especially when a large traffic network must be controlled. The objective of this research is therefore the development of a new coordinated control method for large traffic networks that can optimise the network performance in real time while representing the traffic situation as realistically as possible.

To meet this research objective the control method must be able to optimise the network performance in real time. This means that the computation complexity of the control method must be as low as possible while there will be aimed at optimising the network performance. Literature shows that the optimal solution for a whole network can be reached when coordinated and integrated control is performed, when feedback is used to check the correctness of the solution and when all network effects of traffic are adopted in the control method. Since the only feedback control approaches, optimal; control and model predictive control (MPC), have the big disadvantage of high computation complexity, the real time control of large networks is often not possible. But a solution has been found for this problem: the computation complexity of MPC can be reduced by simplification of the control problem and transforming the non-linear (NLP) control problem into a mixed integer linear programming (MILP) problem. This approach is applied to urban traffic networks but not yet developed for the coordinated and integrated control of large freeway networks. Therefore this research has developed a control method that can be applied to freeway networks and its advantage over other control methods is determined.

Figure 0.1: Control loop of freeway networks
Figure 0.1 shows the total control loop of a freeway network, where the DTM measures in the real traffic network are determined by the MPC controller. The bottom part is developed in this study and consists of two components: at first the traffic model that represents the traffic network, and the optimisation problem that needs to be solved. As stated before, the computation time can be reduced when the control problem is simplified. This means that the traffic model that is used must be as simple as possible, although it is desired to represent all network effects that can occur. Also, the transforming the non-linear optimisation problem into a MILP problem can reduce the computation time. Both steps are performed in this research.

The Piecewise-Affine Network (PAN) model is developed to represent the traffic situation as simple as possible, including the capacity drop and blocking back effects. These two network effects can influence the flows in the network and should be included in the traffic model to represent the traffic situation realistically and to solve or prevent these issues with the control method. Since not all equations of the PAN model are linear, these can be rewritten by the use of logic rules, to transform the PAN model in a Mixed Logical Dynamical (MLD) model. A MLD model consists of linear equations and a set of constraints representing real variables as well as auxiliary and binary variables, and can be used to formulate a MILP problem.

The cplex toolbox (for the MatLab interface) of Tomlab is used to solve the MILP problem for a couple of small networks. This has shown that the control method returns solutions that are logic valid, which means that the solution can be justified analytically, although not all the network effects are represented by the PAN. The computation time needed to solve the control problem of these small networks is very low, which means that it will be possible to control larger networks. The exact scope of the control method could not be determined by these cases, and therefore literature is used to deepen into the factors that influence the computation complexity of MILP problems. This indicates that the computation complexity of the control problem is exponential in worst case and depends on the number of binary variables and constraints. This means that advantage in CPU time, over non-linear control methods, can be gained when a network can be represented by few non-linear equations. Based on this information there is stated that the network size to which the control method can be applied depends on the network lay-out, and this also depends the possible advantage in CPU time over other methods.

The result of this research is a control method that can be used to find the optimal controller settings for DTM measures, and is faster than other network-wide control methods that aim at the optimal solution. Since the computation complexity of the new control method is lower than when traditional MPC is used, larger networks can be controlled in real time.

Key words: large freeway networks, network-wide control, DTM measures, MLD, MILP, efficient MPC, computation complexity
# Table of Contents

Preface ............................................................................................................................... v
Summary ................................................................................................................................. vi
List of figures and tables ....................................................................................................... xii
List of symbols ...................................................................................................................... xiv
  General ............................................................................................................................... xiv
  Original PAN model ......................................................................................................... xiv
  Extended PAN model ....................................................................................................... xiv
  LTM ..................................................................................................................................... xv

1. Introduction ......................................................................................................................... 1
  1.1 Problem definition ........................................................................................................... 1
  1.2 Objective and research questions .................................................................................... 2
  1.3 Contribution of the research .......................................................................................... 3
  1.4 Thesis outline ................................................................................................................ 3

2. Literature study .................................................................................................................... 4
  2.1 Introduction and questions ............................................................................................ 4
  2.2 Network effects in traffic .............................................................................................. 5
    2.2.1 Capacity drop ........................................................................................................... 6
    2.2.2 Shock waves ........................................................................................................... 7
    2.2.3 Blocking back effects ............................................................................................. 7
    2.2.4 Choice behaviour ................................................................................................... 8
  2.3 Dynamic Traffic Management measures .......................................................................... 8
    2.3.1 Dynamic Route Guidance .................................................................................... 9
    2.3.2 Ramp metering ....................................................................................................... 9
    2.3.3 Variable Speed Limits .......................................................................................... 9
    2.3.4 Lane control .......................................................................................................... 9
  2.4 Approximation of optimum .......................................................................................... 10
    2.4.1 Feedback control ................................................................................................... 10
    2.4.2 Coordinated control ............................................................................................... 11
    2.4.3 Integrated control .................................................................................................. 11
  2.5 Network-wide control approach .................................................................................... 11
    2.5.1 Optimal control ...................................................................................................... 11
    2.5.2 Model Predictive Control .................................................................................... 12
    2.5.3 Rule-based control ............................................................................................... 13
    2.5.4 Case-based control .............................................................................................. 13
LIST OF FIGURES AND TABLES

Figure 0.1: Control loop of freeway networks ......................................................... vi
Figure 1.1: Control loop traffic network .................................................................. 1
Figure 1.2: Report structure .................................................................................... 3
Figure 2.1: Specified control loop ......................................................................... 4
Figure 2.2: Control features ................................................................................... 5
Figure 2.3: Bottleneck with capacity drop ............................................................... 6
Figure 2.4: Derivation of shock waves (Hoogendoorn S. P., 2010) ....................... 7
Figure 2.5: Queue formation and disappearing ...................................................... 7
Figure 2.6: Blockage off-ramp due to spillback ....................................................... 8
Figure 2.7: The block diagram of a feed forward control scheme ......................... 10
Figure 2.8: The block diagram of a feedback control scheme ............................... 10
Figure 2.9: Comparison CPU intersection optimisation (Lin, 2011) .................... 12
Figure 2.10: Comparison CPU route optimisation ............................................... 13
Figure 2.11: Overview requirements control method ............................................. 18
Figure 3.1: Set-up control method ....................................................................... 20
Figure 3.2: Schematized overview methodology .................................................. 22
Figure 4.1: The network elements ...................................................................... 23
Figure 4.2: The relations between inflow and outflow of a flow-limiting control measure and of a bottleneck ........................................................................ 25
Figure 4.3: Piecewise constant demand function .................................................. 27
Figure 4.4: Lay out network .................................................................................. 28
Figure 4.5: Demand/inflow link 1 .......................................................................... 28
Figure 4.6: Demand and inflow link 2 ................................................................. 29
Figure 4.7: Demand link 3 .................................................................................... 29
Figure 4.8: Network lay-out blocking back effects ............................................... 30
Figure 4.9: Queue lay-out example 4.2 ................................................................. 31
Figure 4.10: Relation inflow and demand due to spillback ................................... 31
Figure 4.11: Fundamental diagram ................................................................. 34
Figure 4.12: Lay out simple network for quick validity check .............................. 34
Figure 4.13: Overview flow transmission in PAN model ...................................... 35
Figure 4.14: Queue development in PAN model .................................................. 35
Figure 5.1: Schematization of the total control loop ............................................. 38
Figure 6.1: Lay out trivial network ...................................................................... 46
Figure 6.2: Queue development ................................................................. 47
Figure 6.3: Queue development with active bottleneck at start ......................... 48
Figure 6.4: Lay-out non trivial network ............................................................... 49
Figure 6.5: Blocking back effects ...................................................................... 50
Figure 6.6: Queue development when preventing blocking back effects .......... 50
Figure 6.7: Queue development when preventing spillback to underlying network ........................................................................................................... 51
Figure A.1: Example ramp metering ................................................................. 62
Figure A.2: Fundamental diagram example 1 ....................................................... 62
Figure A.3: Example DRIP ................................................................. 63
Figure A.4: Example DSL ................................................................................. 64
Figure A.5: Fundamental diagram of example 3, with a lower speed limit ........ 64
Figure A.6: Example coordinated use DSL and ramp metering ......................... 65
Figure A.7: Example coordinated use DRIP and ramp metering ....................... 65
Figure A.8: Coupled bottlenecks on corridor

Figure A.9: Coupled bottlenecks on two corridors

Figure B.1: Effects of capacity drop on the shockwave derivation

Figure B.2: Derivation shockwaves

Table 2.1: Model characteristics

Table 3.1: Overview theoretical models

Table 4.1: PWC demand pattern

Table 4.2: PWC pattern for the network of figure 4.12

Table 5.1: Truth table (T= true, F=false)

Table 6.1: Piecewise constant demand pattern

Table 6.2: Traffic demand validation network

Table 6.3: Comparison total time spent in the network (h)

Table 6.4: CPU time and number of variables/constraints for cases

Table D.1: Output MILP optimisation
# List of Symbols

## General

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Parameter</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q )</td>
<td>Flow</td>
<td>veh/h</td>
</tr>
<tr>
<td>( q_{cap} )</td>
<td>Maximum flow (at capacity)</td>
<td>veh/h</td>
</tr>
<tr>
<td>( \rho )</td>
<td>Density</td>
<td>veh/km</td>
</tr>
<tr>
<td>( \rho_{jam,x} )</td>
<td>Jam density link ( x )</td>
<td>veh/km</td>
</tr>
<tr>
<td>( \rho_{crit,x} )</td>
<td>Critical density link ( x ) (at capacity)</td>
<td>veh/km</td>
</tr>
<tr>
<td>( L_x )</td>
<td>Length link ( x )</td>
<td>km</td>
</tr>
<tr>
<td>( v )</td>
<td>Speed</td>
<td>km/h</td>
</tr>
<tr>
<td>( v_f )</td>
<td>Free flow speed</td>
<td>km/h</td>
</tr>
<tr>
<td>( N_x(k) )</td>
<td>Number of vehicles on link ( x ) at time ( k )</td>
<td>veh</td>
</tr>
<tr>
<td>( \rho_x(k) )</td>
<td>Density on link ( x ) at time ( k )</td>
<td>veh/km</td>
</tr>
<tr>
<td>( q_x(k) )</td>
<td>Flow on link ( x ) at time ( k )</td>
<td>veh/h</td>
</tr>
<tr>
<td>( v_x(k) )</td>
<td>Speed on link ( x ) at time ( k )</td>
<td>km/h</td>
</tr>
</tbody>
</table>

## Original PAN model

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Parameter</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>( O )</td>
<td>Origin</td>
<td></td>
</tr>
<tr>
<td>( q_o )</td>
<td>Inflow at origin</td>
<td>veh/h</td>
</tr>
<tr>
<td>( d )</td>
<td>Destination</td>
<td></td>
</tr>
<tr>
<td>( q_d )</td>
<td>Outflow at destination</td>
<td>veh/h</td>
</tr>
<tr>
<td>( n )</td>
<td>Node</td>
<td></td>
</tr>
<tr>
<td>( q_{n,in} )</td>
<td>Inflow at node</td>
<td>veh/h</td>
</tr>
<tr>
<td>( q_{n,out} )</td>
<td>Outflow of node</td>
<td>veh/h</td>
</tr>
<tr>
<td>( q_{c,in} )</td>
<td>Inflow at controller</td>
<td>veh/h</td>
</tr>
<tr>
<td>( q_{c,out} )</td>
<td>Outflow of controller</td>
<td>veh/h</td>
</tr>
<tr>
<td>( q_{b,cap} )</td>
<td>Capacity bottleneck</td>
<td>veh/h</td>
</tr>
<tr>
<td>( q_{b,ach} )</td>
<td>Discharge rate bottleneck</td>
<td>veh/h</td>
</tr>
<tr>
<td>( u_c )</td>
<td>Settings controller</td>
<td>veh/h</td>
</tr>
<tr>
<td>( w_c(k) )</td>
<td>Queue length before active controller</td>
<td>veh</td>
</tr>
<tr>
<td>( w_b(k) )</td>
<td>Queue length before active bottleneck</td>
<td>veh</td>
</tr>
<tr>
<td>( t(k) )</td>
<td>Duration stime step ( k )</td>
<td>h</td>
</tr>
<tr>
<td>( N(k) )</td>
<td>Sum of all the queues</td>
<td>veh</td>
</tr>
</tbody>
</table>

## Extended PAN model

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Parameter</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D_k )</td>
<td>Inflow during period ( k )</td>
<td>veh/h</td>
</tr>
<tr>
<td>( t_k )</td>
<td>Duration period ( k )</td>
<td>h</td>
</tr>
<tr>
<td>( q_{c,s} )</td>
<td>Settings controller at end of link ( x )</td>
<td>veh/h</td>
</tr>
<tr>
<td>( S_x(k) )</td>
<td>Sending flow link ( x )</td>
<td>veh</td>
</tr>
<tr>
<td>( O_x(k) )</td>
<td>Outflow link ( x )</td>
<td>veh</td>
</tr>
<tr>
<td>( l_x(k) )</td>
<td>Inflow link ( x )</td>
<td>veh</td>
</tr>
<tr>
<td>( L_x )</td>
<td>Length link ( x )</td>
<td>km</td>
</tr>
<tr>
<td>( \beta )</td>
<td>Turn rate</td>
<td></td>
</tr>
<tr>
<td>Symbol</td>
<td>Parameter</td>
<td>Unit</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>$q_{M,i}$</td>
<td>Capacity link $i$</td>
<td>veh/h</td>
</tr>
<tr>
<td>$q_{dch}$</td>
<td>Discharge rate (capacity drop)</td>
<td>veh/h</td>
</tr>
<tr>
<td>$t$</td>
<td>Time</td>
<td>h</td>
</tr>
<tr>
<td>$\Delta t$</td>
<td>Time difference</td>
<td>h</td>
</tr>
<tr>
<td>$N$</td>
<td>Cumulative vehicle number</td>
<td>veh</td>
</tr>
<tr>
<td>$N(x^0_i, t)$</td>
<td>Cumulative vehicle number at begin of link $i$</td>
<td>veh</td>
</tr>
<tr>
<td>$N(x^t_i, t)$</td>
<td>Cumulative vehicle number at end of link $i$</td>
<td>veh</td>
</tr>
<tr>
<td>$S_i(t)$</td>
<td>Sending flow link $i$ (between $t$ and $t + \Delta t$)</td>
<td>veh</td>
</tr>
<tr>
<td>$R_j(t)$</td>
<td>Receiving flow link $j$ (between $t$ and $t + \Delta t$)</td>
<td>veh</td>
</tr>
<tr>
<td>$w_j(t)$</td>
<td>Shockwave speed</td>
<td>km/h</td>
</tr>
<tr>
<td>$q_{measure}(t)$</td>
<td>Outflow traffic control measure</td>
<td>veh/h</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

Congestion on freeway networks is a major problem in many countries. It causes delays for the road users, mostly expressed in lost vehicle hours. Governments want to reduce these lost vehicle hours, which can be done by minimising or changing the traffic demand, road expansion, encouraging public transport use and traffic management (Kennisinstituut voor Mobiliteit, 2009). The last years, a strong focus has been developed on sustainable and dynamic traffic management (DTM) since expansion of the roads is expensive and not always possible due to the available free space and environmental and nuisance restrictions.

This thesis will research the network-wide control of DTM measures to improve the overall performance of traffic networks. Paragraph 1 of this chapter will elaborate on the challenges and problems regarding this subject, which will lead to the research objective of this thesis in the next section.

1.1 PROBLEM DEFINITION

To improve the performance of a freeway network DTM measures can be used. There is a variety of DTM measures available to guide and control the vehicles in the network, such as route guidance, ramp metering and lane management. These measures must be controlled: a control method can determine the settings for the measures that will maximize the network performance.

Figure 1.1 shows the control loop of a traffic network, which consists of 5 parts. The traffic demand is the external input of the traffic network, which influences the network performance. This is measured by sensors and gives input to the control method.

In the control method, DTM measures are used as instruments to influence the situation in the network (the traffic state). When a closed loop approach is used, the effects of the measures on the network performance give feedback to the control method which adjusts the DTM measures when needed.

Currently in traffic management, many DTM measures are controlled on local level to solve a single bottleneck and the effectsen jij of these measures on the whole freeway network are mostly not taken into account. This can result in solving a single bottleneck but it can cause decrease the performance of the network as a whole (Papageorgiou, Hadj-Salem, & Blosseville, 1991). To prevent this from happening and to maximize the network performance, a network-wide approach for the control of the DTM measures is needed. One of the main problems of network-wide control approaches is the computational complexity of the controllers in the network (Lin, 2011). The computation time needed to determine the controller settings increases when expanding the network size and this can result in computation times which are larger than real time, which makes real time control impossible.
When it is desired to optimise the network performance of large scale networks (for example controlling the whole Randstad instead of only a freeway corridor), real time control can be a problem due to the computational complexity. Since real time control is needed to respond to the actual traffic situation, a way to handle this problem must be found. Options are the use of scenarios instead of determining the optimum (Wismans, Bliemer, & van Berkum, 2010), to implement only one type of control measure (Kotsialos & Papageorgiou, 2004), to estimate the optimum (Sedek, Demetsky, & Smith, 1999) or to simplify the traffic network (Hegyi, De Schutter, & Hellendoorn, 2005). None of these control methods reaches the true optimum which leaves possibilities to improve the network performance.

1.2 **OBJECTIVE AND RESEARCH QUESTIONS**

This thesis studies the possibilities of developing a new control method that can handle the computation complexity of large freeway networks better than current methods. Therefore the following research objective is formulated:

*To develop a control method for large traffic networks that can optimise the network performance in real time.*

When designing a control method, the scheme of figure 1.1 must be further developed. At first, the traffic demand must be specified. Here, two components are of importance: the dynamics of the demand and the distribution of the demand based on the origin-destination matrix over the routes in the network. In reality the demand has a dynamic nature and road users can change their route choice based on the traffic situation, but a static demand function and fixed routes can also be chosen to simplify the real situation. Next, the traffic network must be represented by a traffic model to predict the situation in the network. Many mathematical traffic models are available which differ in the network effects (such as blocking back and shockwaves) they represent, the DTM measures that can be implemented and whether they can digest static or dynamic demand. A cost function is used to determine the network performance $J$, which gives input to the control system. Also for the control, many approaches are available, depending on the control features and structure and the DTM measures that will be implemented (Lin, 2011).

To meet the research objective choices have to be made about the components of the described control loop before the method can be developed. Therefore the following research questions are formulated:

1. *What are the requirements for a real time network-wide control method?*
2. *How can the computation time be reduced?*
3. *What is the value of the new control method, compared with other control methods in terms of computation time and approximation of the optimal solution?*

The first question will deliver the requirements and wishes for the control method: which control features must be included, which control features are wanted to be included? Based on the requirements, there will be looked into options to reduce the computation time of control methods as mentioned in question 2. Together this will determine the content of the control method. When this method is designed, the value of this method must be determined (question 3) by verification and by comparing the method in terms of computation time and approximation of the optimal solution.

These questions will be answered in separate sections, as described in the thesis outline.
1.3 CONTRIBUTION OF THE RESEARCH

The development of a new network-wide control method with a lower computation complexity can be used for the real time control of larger traffic networks than possible with the current methods. This thesis will demonstrate the value of network-wide control, and a method that can control large traffic networks in real time will be able to improve the network performance. In practice this can lead to the coordinated, real time control of large traffic networks, which will reduce the lost vehicles hours on the freeways.

From academic perspective, this thesis will be at first a good start for the actual development of a method that can be applied in practice. Further, the research can be used in academic studies after the effects or the possibilities of network-wide dynamic traffic management, where those studies are now limited by the computation time of the control methods.

1.4 THESIS OUTLINE

This document starts with a literature study in chapter 2 to gain more information about the components of a control method and the strengths and weaknesses of the current control approaches. This information leads to a set of requirements and wishes for the control method, which will answer the first research question.

Based on the requirements and wishes for the control method, chapter 3 will set up a methodology for the research in the third chapter. There will be decided in which way the computation time of a control method can be reduced (third research question). This leads to a description of the control and modelling approach and the way in which the last research question can be answered. Chapter 4 describes the traffic network model that is used in this research and the extensions made to fit the requirements as formulated in the conclusion of the literature study and the methodology. The control method, linked to this traffic model, is developed in chapter 5 and will be verified and validated in the following chapter. Here, the last research question will be answered.

Chapter 7 presents first the conclusions of the research and will indicate in which way the research objective is achieved. The strengths and weaknesses of the new control method are given in the reflection section and will lead to a set of recommendations for further research.
2. LITERATURE STUDY

To meet the objective as stated in the previous chapter (to develop a real time control method for large traffic networks), more information is needed about the options to control traffic networks. The goal of this literature study is to obtain this information and use it to construct the approach for this research.

The chapter will start with elaborating on the kind of information that is needed to develop a new control method, which results in a few questions for this literature study. These questions will be answered in the subsequent sections. At the last section, the answers will be combined to develop the approach for this research.

2.1 INTRODUCTION AND QUESTIONS

The control loop as pictured in figure 1.1 indicated how a traffic network can be controlled, but since the controller part is not specified yet this will be done here. Figure 2.1 shows the specified controller loop. The controller consists of a traffic model which represents the situation in the traffic network and a control method which determines the settings for the DTM measures based on the network performance given by the traffic model.

The total of the bottom part of the figure above is indicated as “controller”, but should not be mistaken for the physical controllers: the dynamic traffic management measures. This part must be developed to meet the research objective, what means that there must be given substance to the following 5 components:

1. Traffic demand
2. Traffic model
3. Traffic state (measured by sensors)
4. Control method
5. DTM measures (controllers)

The choice for a traffic model and the choice for a control approach influence the network performance and the way in which the DTM measures are controlled. Therefore, it is important to research the features of traffic models and control approached and their effects on the network performance and computation time. This information is needed when deciding how to design the new control method.
Figure 2.2 shows the features which are of importance when choosing a traffic model and a control approach. In the research objective there is stated that real time control is required and that the large networks must be served, this is indicated in figure 2.2 by the red square. The way in which the other features are served is not stated yet and will be determined by the help of literature: the network effects that are used to describe the traffic network, the DTM measures that can be implemented, the choice between a static or dynamic approach and the extent to which the optimum is reached.

![Diagram of traffic control system]

Figure 2.2: Control features

To determine how these features must be served by the control method and the traffic model, literature is used to obtain more information about these aspects.

For these six subjects, the following research questions are formulated for the literature study:

1. Which network effects can be identified, when do they occur and how large are they?
2. Which DTM measures can be identified and what can be the effects on the network performance?
3. Which control features are of importance to reach the optimal control solution?
4. What are the strengths and weaknesses of the current control approaches?
5. What are the strengths and weaknesses of a static versus dynamic approach?
6. Which traffic models can be used to describe large freeway networks?

Each of these questions will be discussed in separate sections and will contribute to the conclusion, where the research questions are answered and requirements and wishes are set for the control method.

2.2 Network effects in traffic

Network effects are traffic phenomena which can influence the situation in a network in terms of flows, speed and density negatively. When a traffic network is controlled, the occurrence of these effects must be taken into account: preventing these effects from happening can improve the network performance.

Literature identifies the following four network effects in traffic networks (Hoogendoorn, Taale, Wilmink, van Katwijk, Immers, & Schuurman, 2011):

1. Capacity drop at active bottlenecks
2. Shockwaves
3. Blocking back effects
4. Sub optimal choice behaviour of road users
When these effects occur and how they influence the network performance is of importance for designing the control method. Which effects should be included in the control method will be determined based on the degree in which the influence the network performance. Therefore, each effect will be described in the following sections.

### 2.2.1 Capacity Drop

The capacity drop can be described as the phenomenon that the discharge of an active bottleneck \( q_{dcb} \) is lower than the free flow capacity \( q_{cap} \) of the bottleneck (Cassidy & Bertini, 1999). An bottleneck will be(come) active when at least 1 of the following situations occurs:

- A queue is presented at the link(s) upstream the bottleneck
- The flow and/or density is higher than the capacity of the bottleneck

The effects of a capacity drop on the network performance are identified in several studies:

1. Chung et al. (2007) give 15% at merge bottlenecks, 6% at lane reduction and about 5% at horizontal curves
2. Hall and Agyemang-Duah (1991) give 6% at merge bottlenecks
3. Brilon and Zurlinden (2003) give 24% at merge bottlenecks

These numbers only give the reduced outflow of an active bottleneck, but do not take the effects upstream into account. To illustrate the effects of a capacity drop a simple example is given.

**Example 2.1**

Consider a freeway stretch with an origin providing the traffic demand \( q_{in} \), a main road with an off-ramp and on-ramp and a destination \( D_2 \) at the end of the stretch, as shown in figure 2.3. In this example, the following parameters are considered:

\[
\begin{align*}
q_c &= \text{capacity at free flow} = 2000 \text{ vehicles/hour} \\
q_{dcb} &= \text{capacity in congested state (discharge capacity)} = 1800 \text{ vehicles/hour} \\
q_{in} &= \text{traffic demand} = 2000 \text{ vehicles/hour}
\end{align*}
\]

When the bottleneck is activated, it will not be able to accommodate the total traffic demand. This causes the formation of a queue upstream that can block the off-ramp and a higher density that can cause shockwaves. When the bottleneck is deactivated the traffic demand can pass in free flow, and queues will not be formed. However, the outflow of the bottleneck will be higher than in active state which influences the traffic situation downstream since the traffic demand will be higher at the next bottleneck.

![Figure 2.3: Bottleneck with capacity drop](image)

Measurements show that the capacity drop occurs when the density at a point is higher than the critical density in free flow (Chung, Rudjanakanoknad, & Cassidy, 2007), but in studies sometimes the traffic demand in relation to the capacity in free flow is used to determine the presence of a capacity drop (Hegyi, De Schutter, & Hellendoorn, 2005). The different approaches (density versus flow) for the representations of the capacity drop will be discussed further when there is decided to construct a traffic model which adopts the capacity drop.
2.2.2 Shock waves

When the traffic state in a network changes, for example due to a higher traffic demand or a blockage on the road, shockwaves refer to the boundary between the two states (Segl, 2008). The shockwave speed indicates the speed with which the head and the tail of a congested state move upstream (Hoogendoorn S. P., 2010). Figure 2.4 shows how the formation of a shockwave can be derived from the fundamental diagram.

![Figure 2.4: Derivation of shock waves (Hoogendoorn S. P., 2010)](image)

In this fundamental diagram, four different states are identified due to a temporary blockage:

1. Uncongested state with free flow speed
2. Standstill, where the flow is zero (due to the blockage)
3. An empty network, where the flow and density are zero (downstream the blockage)
4. Full utilisation of the capacity at the critical density (when the blockage is solved)

The presence of shockwaves can reduce the outflow of the network by 30% on average (Hegyi, Hoogendoorn, Schreuder, Stoelhorst, & Viti, 2008). There are two ways to deal with shockwaves: preventing and resolving. In the first case, disturbances in the network are prevented so shockwaves will not occur. The other option is to resolve the shockwave by the activation of dynamic speed limits whereby the speed is limited. Since the density stays the same, the flow must decrease in this situation (according to the basic equations for the fundamental diagram $q=kv$). After some time, the shock waves will disappear (Hegyi, Hoogendoorn, Schreuder, Stoelhorst, & Viti, 2008).

2.2.3 Blocking back effects

When the inflow at a location of bottleneck is higher than the capacity a queue is built up, which is called spillback of congestion. Figure 2.5 shows the number of vehicles in a queue when a bottleneck has a capacity of 1200 vehicles and the inflow is determined by:

$$q_{in}(t) = \begin{cases} 
2400 \text{ vehicles/hour} & t \leq 20 \text{ min} \\
600 \text{ vehicles/hour} & t > 20 \text{ min}
\end{cases}$$

![Figure 2.5: Queue formation and resolution](image)
When the queue length exceeds the distance between the bottleneck and an off-ramp upstream, this off-ramp is blocked what has consequences upstream: the turning vehicles cannot leave the link (figure 2.6). The effects of this phenomenon on the network performance are called “blocking back effects”.

Figure 2.6: Blockage off-ramp due to spillback

Blocking back effects can only be identified by a model when it assumes horizontal queuing, which is not the case in every traffic model. Previous research has proven the importance of modelling blocking back effects when assessing the effects of DTM: when it is not taken into account, the travel times in the network are underestimated as well as the effects of route information systems (Knoop, van Zuylen, & Hoogendoorn, 2008).

2.2.4 Choice behaviour

Road users make choices about their travels based on incomplete information and with the aim to maximise their own benefit, regardless the effects this will have on the other road users (Hoogendoorn, Taale, Wilmink, van Katwijk, Immers, & Schuurman, 2011). In the traffic network this will deteriorate the situation since the road users pursue their own benefits instead of aiming at maximisation of the network performance. This phenomenon is mostly applicable on departure times and route choice and causes non-optimal usage of the capacity in a network.

When the traffic demand in the network exceeds the capacity, benefits can be achieved by guiding the road users over time and space. (Zhou & Wu, 2006). This will improve the distribution of traffic over the network and can decrease the traffic demand at bottlenecks.

2.3 Dynamic Traffic Management Measures

The control of the freeway network is done by implementing dynamic traffic management measures. To determine which measures will be used in the control method, this paragraph describes a couple of them and the extent to which they can improve of the traffic performance.

A large variety of DTM measures can be identified in literature, not only be to control traffic flow but also to guide and to inform (van Kooten & Adams, 2010). For this research only the first group is of importance, since the objective is controlling the traffic network. Further, only measures for freeway networks are relevant. Van Zuylen (2002) identifies the following options for traffic management at freeways, which will be discussed in more detail in this paragraph:

- Dynamic route guidance systems
- Ramp metering
- Variable speed limits
- Lane control
2.3.1 Dynamic Route Guidance

By guiding vehicles over the network, the existing capacity is better utilised (Zuurbier, 2010). Many theoretical studies and practical tests have shown the potential success of Dynamic Route Guidance (DRG). Application in Germany shows a travel time advantage of over 5% in the whole network (Helling & Schoenharting, 2006) and the modelling study of Kotsialos et al (2002) even gives an improvement of 29%.

2.3.2 Ramp metering

Ramp metering limits the flow from an on-ramp entering the main road. This can serve two objectives: preventing spillback towards an off-ramp upstream the main road (Papageorgiou & Kotsialos, 2002), and preventing the activation or stimulate the deactivation of a bottleneck (Hegyi, 2004). The extent to which ramp metering can contribute to these objectives depends on the traffic situation and the constraints such as the maximum queue length on the on-ramp. Tests with different approaches to control the ramp metering on local or network level give good results: a decrease of the total travel time by on average 5% in the research of Hegyi (2004) and a decrease of 16% in the study of Kotsialos et al (2002).

2.3.3 Variable Speed Limits

Variable Speed Limits (VSL) can be used for two goals: suppressing shockwaves and decreasing the outflow of a section. The success of the first function is proven by testing the SPECIALIST algorithm for VSL at the Dutch A12 freeway: 55% till 83% of the shockwaves were resolved (TU Delft, 2011).

The effectiveness of VSL in the aim for outflow limitation is less clear. Evaluations of practical tests with lower speed limits in the Netherlands show that this does not result in a better network performance in terms of total travel time or total outflow (Drewes & Vink, 2011). Other studies show the theoretical success of VSL: Hegyi (2004) uses VSL to limit the inflow from the mainstream which can prevent or resolve the activation of bottlenecks. The study by Papageorgiou et al (2008) concludes that VSL are only successful when implemented just before a bottleneck, to limit the inflow at this bottleneck without reducing the outflow at the bottleneck upstream. This application of VSL can be compared with mainstream control, since it is meant to limit the outflow and spillback can occur when the inflow of a section is higher than the outflow restricted by the measure.

2.3.4 Lane Control

The whole of measures that influence the availability of lanes is called lane control and includes peak hour lanes, lane closure and dynamic lane markings (Papageorgiou, Diakaki, Dinopoulou, Kotsialos, & Wang, 2003). The dynamic availability of lanes can be matched to the traffic demand, but can also be tailored to a bottleneck: the traffic flow into a bottleneck can be decreased by lane closure upstream, or the traffic outflow of a bottleneck can be secured by opening an extra lane downstream.

Evaluation studies of peak lanes in the Netherlands show the success of this measure: the congestion weight did decrease with over 50% on the A12 (Wilmink, Vonk, & Taale, 2010) and the spillback on the A27 almost totally disappeared due to the peak lane (AVV, 2000). But there must be stated that peak lanes cannot be implemented on all trajectories, since this depends on the available space. Dynamic lane marks can deal with locations where the space is limited: the same width is used for $x$ normal lanes or $x+1$ narrowed lanes during peak hours. The deviation of the available space is done by LED lighting, and adds an extra $\frac{100}{x}$% to the capacity (Heymans, 2005). This measure is not implemented yet in the Netherlands due to the safety issues and the technical requirements.
2.4 **Approximation of optimum**

The research will focus on maximising the extent to which the optimal solution is reached since this will improve the network performance most.

The extent to which the optimal solution is reached can be determined by percentages in which the solution given by the control method matches the optimal solution. This means that this can only be determined when the optimal solution is known but based on cases where this optimum was known, the factors that influence the approximation of this optimum could be identified: in a network it depends on the presence of feedback control and whether coordinated and integrated control strategies are used.

Feedback control means that a closed loop system is used: there is checked whether the predicted optimal control settings were correct, and when needed adjustments are done. Coordinated control strategies consider multiple control measures of the same type that work together in a network, while integrated control strategies consider the combination of different types of control measures (Kotsialos, Papageorgiou, Mangeas, & Haj-Salem, 2002).

Several studies have been performed after feedback control, coordinated control and integrated control and their findings are discussed in the following sections.

2.4.1 **Feedback control**

Figure 2.7 and figure 2.8 show the difference between a system without and with feedback control: when feedback control is applied, additional measurements are done to check the effects of the controller on the process (Franklin, Powell, & Enami-Naeini, 2002). In these control schemes the traffic behaviour is predicted by the control method, but the traffic state in the real network can also be influenced by unpredicted disturbances.

![Figure 2.7: The block diagram of a feed forward control scheme](image)

![Figure 2.8: The block diagram of a feedback control scheme](image)

Feedback control often results in a quicker and better performing system, and may stabilize unstable processes (traffic is unstable in congestion), but may also destabilise stable processes (if not properly designed). In addition, while feed forward controllers can only react on measurable disturbances, feedback controllers can also react on immeasurable disturbances if the effects of the disturbances are observable through the system output (van Lint, van Zuylen, Hegyi, Hoogendoorn, Bliemer, & Pel, 2009). There can be concluded that well designed feedback control systems can lead to better network performances that feed forward control systems.
2.4.2 Coordinated control

Coordinated control means that the measures of the same type are coupled to each other in the network. Unlike local control, the controllers work together to maximize the performance of the whole network. This can be expressed by a measure creating congestion at one point to improve the performance overall.

Several studies after coordinated control (Kotsialos & Papageorgiou (2001) and Lin et al (2010)) use dynamic traffic models to determine the delay between measures. When these models use a correct queuing method, the correctness of the delay function is guaranteed. The effects of coordinated control compared with local control are not often quantified in literature, but Yuan (2008) gives an improvement of 2% and Kotsialos & Papageorgiou (2001) give 15%.

2.4.3 Integrated control

The study of Hegyi (2004) has shown that it can be beneficial to use more types of DTM measures to control a bottleneck: due to constraints considering queue lengths it is not always possible to deactivate a bottleneck with a single measure, and adding another type of measure will extend the solution space. In the case study of Hegyi (2004), the integration of ramp metering and dynamic speed limits as well as the integration of ramp metering and mainstream control improved the total travel time with about 15% compared to the ‘no control’ case, where the effect of only ramp metering was just 5%.

A study of Kotsialos et al (2002) after the integration of ramp metering and route guidance also shows a big improvement due to the integrated control: the total travel time in the network is decreased with 30 to 36% compared to the ‘no control’ case. Here, the effects of the individual measures are also larger: an improvement of 16% due to ramp metering and 29% due to the route guidance.

These studies prove that integrated control leads to further improvement of the network performance, compared with the use of just one type of measure.

2.5 Network-wide control approach

The scheme of figure 2.2 identifies the network size and the extent to which the optimal solution is approximated as control features. Also, it is stated that the control method must be able to compute the solution in real time: real time control is required, which sets restrictions to the computation complexity. Based on the choices for these features, the most suitable control approach can be determined. For freeway networks the following control approaches are identified (Lin, 2011):

1. Optimal control
2. Model Predictive Control (MPC)
3. Rule-based control
4. Case-based control

The suitability of each approach depends on the way in which they handle the control features. Therefore these control approaches are evaluated on the inclusion of feedback, integrated and coordinated control and the computation time needed to perform the optimisation.

2.5.1 Optimal control

Optimal control is a mathematical approach to determine the control variables given a cost function which has to be minimized or maximized. Optimal control can only be used for real time optimisation when an efficient algorithm is found, which can deal with the non-linearity of the network-wide control of freeway networks
network, otherwise the needed computation time for the optimisation exceeds the running time of
the model. Another disadvantage is the fact that most optimal control strategies are open-loop,
which means that feedback is not included. The ALINEA algorithm does use feedback control, but is a
local control strategy and not suitable for network-wide control.

The Advanced Motorway Optimal Control (AMOC) strategy is a type of optimal control which is used
to optimise the network performance by the use of ramp metering or route guidance (Kotsialos &
Papageorgiou, 2004). It is implemented in the METANET model and has given good results for the
case ring road Amsterdam.

2.5.2 Model Predictive Control

Model Predictive Control (MPC) does use feedback: the optimal control is repeated in a rolling
horizon way, which makes the approach closed-loop. This means that, in theory, a MPC approach
can maximize the outflow of the whole network better than when optimal control is applied.
However, the computation complexity is a big challenge for the real time control of traffic networks.

To deal with the large computation time of MPC, different studies have approximated the control
problem and recast it as a mixed integer linear programming (MILP) problem which is much easier
(and thus faster) to solve. The study of Lin (2011) proposes an approach for reducing the
computation time needed that consists of two steps:

1. Simplification of the traffic model
2. Reformulating the non-linear programming (NLP) problem into a MILP problem

This approach is applied to a couple of traffic related studies. It is used to optimise the control of
intersections in urban areas (Lin, 2011) and to optimise the route choice in freeway networks (van
den Berg, De Schutter, Hegyi, & Hellendoorn, 2008). The results of these studies are recalled here to
get insight in the difference in computation complexity between a MILP problem and non-linear
control methods and the effects the simplifications can have on the solution.

![Figure 2.9: Comparison CPU time intersection optimisation (Lin, 2011)](image)

Figure 2.9 shows the CPU (central processing unit) time needed to solve the optimisation problem
for the control of an urban network. A network with four coupled intersections was optimised with
sequential quadric programming (SQP) and with mixed integer linear programming (MILP) for
different scenario’s concerning the traffic demand. The computation time needed to solve the
problem differs much between these two approaches: the optimisation through MILP is 100 to 400
times faster. The solution of the MILP optimisation approximates the solution of SQP: in the first
case the solutions match, in the other cases the variation is 2% to 8%.
Figure 2.10: Comparison CPU route optimisation

The same approach is also used in the study after route choice: a network with an origin and a destination, connected by two routes is optimised with route choice as controller (van den Berg, De Schutter, Hegyi, & Hellendoorn, 2009). The network is optimised for four different optimisation algorithms, including the use of and MPC through MILP control approach. All these algorithms gave the same solution, but the time needed to achieve this solution differed a lot, as displayed in figure 2.10.

2.5.3 RULE-BASED CONTROL

The rule-based control of freeways makes real time decisions based on rules that are basically formulated as “if X happens, than Y”. This approach is often applied for the coordination of local control methods.

An example of rule-based control of traffic networks is the HERO algorithm, which adopts the control algorithm ALINEA for ramp metering. ALINEA controls the ramp-metering on local level, by deciding how much vehicles can enter the mainstream based on the flow of the mainstream. The HERO algorithm takes all the local ramp metering systems in a network into consideration to determine which streams must get priority to improve the network performance. Although an optimal control will not be achieved, the method is efficient and fast what makes it well suitable for real-time control or larger size networks (Yuan, 2008).

2.5.4 CASE-BASED CONTROL

Case-based reasoning is based on the observation that when people solve a problem, they often base the solution on one that worked on a similar problem in the past. For the control of traffic networks, this means that a problem does not have to be solved from scratch by the use of a complex mathematical model. Instead there will be looked into similar problems, that solution will be taken and revised when needed.

The big advantage of this method is the easiness and the ability to deal with situations where not all the information is known (Sedek, Demetsky, & Smith, 1999), but it will not determine the optimal network performance.

2.6 STATIC VERSUS DYNAMIC APPROACH

A static or dynamic approach can be chosen for the control method. This means that the model that represents the traffic situation can be static or dynamic in terms of traffic demand and how the traffic behaves. A static approach does not represent a variety of time, while the dynamic does.

The choice between a static and dynamic approach for the traffic demand and the traffic behaviour has consequences for the network effects that can be represented, and the computation complexity of the control method.
2.6.1 Traffic Demand

When a static demand is assumed, the inflow that does not change over time, a dynamic demand represents an inflow that varies over time. Static demands in traffic do not occur in reality, but can be a good assumption when developing a new approach since it simplifies the situation. An example of simplifying the situation by the use of static demands is the study of (Hegyi, De Schutter, & Hellendoorn, 2005) which assumes static demands and gives recommendation for further expansion to dynamic demands. Using dynamic demand instead of static demand has consequences for the control approach: the DTM measures must adapt to the varying traffic demand, which increases the computation complexity since more calculation are needed. There can be stated that a static demand is not realistic, but can be used when simplification of the situation is needed.

2.6.2 Traffic Assignment

For the representation of the traffic behaviour in the network, also a distinction can be made between a static and dynamic approach. Traffic assignment concerns the routes choice of road users, which represents how vehicles will be distributed over the network. Static traffic assignment (STA) uses static travel time functions to distribute the traffic over the routes. This means that the value of the routes does not change over time and the distribution of the traffic over the routes will be the same at all times. Dynamic traffic assignment (DTA) adopts the network effect ‘choice behaviour’ and adjusts the value of the routes depending on the situation in the network, what changes the distribution of vehicles over the routes to create a new optimal situation (Bovy, Bliemer, & van Nes, 2006). Due to all the traffic information services such as the radio, route guidance systems and dynamic route guidance panels road users do have the opportunity to change route because of the traffic situation, what is assumed in DTA. But this is also mathematically more complex, wherefore STA can give a simplified vision.

Besides the absence of route choice behaviour, the location and physical length of congestion is also not reproduced well by STA. The presentation of Bliemer et al (2011) shows that in traditional STA models, the queues are located at the bottlenecks instead of upstream the bottleneck. This gives incorrect travel times in congested situations, due to the delay function between bottlenecks. Realistic DTA models give a correct definition of the queue lengths, congested speeds and shockwaves, and approach the traffic situation in a network therefore more realistic.

2.7 Traffic Models

As mentioned before, the control method will be applied to a traffic model that described the situation in the network. Which models are suitable for this research depends on the network effects that must be represented, the DTM measures that will be applied and the choice for a static or dynamic approach. Next, the model must also meet the requirements that are set by the control approach. The variety of available traffic models can be categorised based on their characteristics as shown in table 2.1.

<table>
<thead>
<tr>
<th>Scope</th>
<th>Microscopic</th>
<th>Mesoscopic</th>
<th>Macroscopic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamics</td>
<td>Static</td>
<td>Dynamic</td>
<td></td>
</tr>
<tr>
<td>Network effects</td>
<td>Capacity drop</td>
<td>Shockwaves</td>
<td>Blocking back effects</td>
</tr>
<tr>
<td>DTM measures</td>
<td>DRG</td>
<td>Ramp metering</td>
<td>VSL</td>
</tr>
</tbody>
</table>

Table 2.1: Model characteristics

At this point, only the scope of the traffic model is determined; the objective of the research indicates that a macroscopic model is needed, since this type of model is suitable for large-scale simulations of traffic flows in networks (Hoogendoorn & Bovy, 2001). The model must be dynamic.
because it is needed to model the changes over time in a network due to the DTM measures. All these requirements limit the number of suitable theoretical models.

Focussing on dynamic macroscopic models, a distinction can be made between first and higher order models. First order models use only one partial differential equation to describe the evolution of the traffic state, while higher order models use more equations (Haut & Bastin, 2005). First order models are sufficient to describe the size and end of the queue, but have been criticized on a few points which led to the development of higher order models (Hoogendoorn S. P., 2010). These models use more equations what gives a more detailed view of the traffic situation, but also adds complexity. Complexity of the equations results in long computational time, what is undesired for the control of the network. Although it is not decided yet which control approach will be used, a fast computation time of the model is of value for every approach.

The studies of Hoogendoorn & Bovy (2001) and Tyagi et al (2009) show overviews of the developed traffic models till that moment. A selection of these models is made, based on suitability and relevance of the models for this research.

2.7.1 LWR

The kinematic wave model of Lighthill and Whitham, with extensions by Richards, is one of the oldest traffic models where the dynamics in a network are approached as fluids. The situation is a network is described by the relations between density ($\rho$), flow ($q$) and speed ($v$): 

$$q = \rho v$$

where an equilibrium between the difference in density over time ($\partial \rho/\partial t$) and the difference in flow over distance ($\partial q/\partial x$) can be formulated (Lebacque & Khoshyaran, 2004):

$$\frac{\partial \rho}{\partial t} + \frac{\partial q}{\partial x} = 0$$

The LWR model adapts the occurrence of shockwaves, the speed ($\omega$) at which a shockwaves propagates upstream is calculated by the difference in flow and density over time and space:

$$\omega = \frac{q_1 - q_2}{k_1 - k_2}$$

This model is very simple, but also has some drawbacks (Kerner, Konhäuser, & Schilke, 1996):

- The kinematic model does not admit deviations from the speed–density $v(x, t) = v^e(r(x, t))$ relation. In other words, drivers react instantaneously to changing traffic conditions.
- Discontinuous solutions result irrespective of the smoothness of the initial solution. This contradicts the observed real-life flow behaviour.
- The LWR-model is unable to capture the formation of localized structures, phantom-jams, hysteresis, and stop–start waves under specific conditions. This is unfortunate, since realistic modelling these phenomena is imperative for describing real-life traffic flow.

Besides the disadvantages as pointed out in the literature, there must also be noted that the capacity drop is not included in the LWR model and the transmission of vehicles between sections is unrealistic. The downstream occupancy is not explicitly used here what means that traffic can be send into the next section even when this section is fully occupied.
2.7.2 CTM

As reaction on the earlier mentioned shortcoming of the LWR model, Dagonzo developed the Cell Transmission Model. This model is a discrete version of the LWR and divides the links in a network in cells of equal length, wherefore the traffic flow \( q \) is not only determined by the speed \( v \) and the density \( \rho \) but is also restricted by the capacity of the link \( q_{\text{max}} \) and the maximum density \( \rho_{\text{jam}} \):

\[
q = \min\{v\rho, q_{\text{max}}, v(\rho_{\text{jam}} - \rho)\}, \quad \text{for } 0 \leq \rho \leq \rho_{\text{jam}} \quad \text{(Daganzo, 2008)}
\]

This means that the outflow of a cell in free flow is, just as in the LWR model, determined by a function of the free flow speed and density. In other states, the outflow is limited by the capacity and the speed with which disturbances in the next section propagate backwards. Although the basic model does not consider a capacity drop when the critical density is reached, several software models, such as Fastlane, have extended the CTM with the capacity drop (Schreiter, van Lint, Hoogendoorn, Muhurdarevic, & Scheerder, 2011). This model is, compared to the LWR model, a more suitable approach for this research.

2.7.3 LTM

In 2005, a new discrete formulation of the kinematic wave model is proposed which is similar to the CTM but substantially smaller in computational complexity: the Link Transmission Model (Yperman, Logghe, Tampere, & Immers, 2005). As the name states, the model calculates the transmission of vehicles between links instead of cells as is the case in the CTM, what gives the gain in computational time. Another advantage of the LTM over the CTM is the possibility to calculate the locations of queues exactly, where the CTM only indicates if a cell is in free flow or congested.

The LTM calculates for each link at every time step \( k \) the possible number of vehicles that can leave the link \( (S_i(k)) \) and the number of vehicles that can be received by the link downstream \( (R_j(k)) \). The sending flow is determined by the number of vehicles that can reach the end of link \( i \) at time \( k \), and is restricted by the capacity of link \( i \). The receiving flow takes the presence of shockwaves into account, the maximum density of link \( j \) and is also restricted by the capacity of the link (Yperman I., 2007). Also in this model the capacity drop is not included, but can be added to the receiving flow of a link in the same way as done in studies with the CTM.

2.7.4 Payne

The second order model of Payne uses two state variables (instead of one as the previous mentioned three models) to describe the traffic states: traffic density and average velocity. This model is able to address the instability of traffic and can therefore identify the occurrence of start-stop waves and localised traffic jams (Hoogendoorn & Bovy, 2001) and also adopts the capacity drop. These factors are major advantages of the Payne model, but the original model indicates wave characteristics which are faster than the mean speed of the simulated traffic. Therefore, Papageorgiou has made improvements to this model twice. This resulted in the Payne-Papageorgiou model (Papageorgiou M., 1983) which is one of the most-used models due to the extensive reproduction of the traffic situation. The consequence of the extensiveness is the big disadvantage of the model: high computational complexity due to the extended equations.

2.7.5 Gas-kinetic

Traffic can also be described according to the dynamics of a fluid of a gas, what gives continuum models. The starting point of the gas-kinetic models is the so called phase-space density (PSD); the density depends on the location, time and the speed: \( k(x, t, v) \) and will change due to convection, acceleration and deceleration of the vehicles (Hoogendoorn S. P., 2010).
This type of models is able to reproduce the velocity-density relation of traffic (just as the first order models) as well as the observed spectrum of non-linear phenomena and their characteristics (Helbing, Hennecke, Shvetsov, & Treiber, 1999). The reproduction of both aspects is the big advantage of this model, but the disadvantage of the gas-kinetic model is its complexity, what makes it hard to derive an appropriate numerical solution.

2.7.6 PAN MODEL

At last, the Piecewise Affine Network model (Hegyi, De Schutter, & Hellendoorn, 2005) is discussed. This static model is designed for the control of traffic networks and describes a network by the presence of origins, destinations, nodes, links, bottlenecks and controllers (DTM measures). The time is divided in stages and to determine the number of vehicles in the network (or on the links) at every stage, not the fundamental diagram is used but only the flow (q). The capacity drop is included in the model, but a static demand is used and blocking back effects are not taken into account. Also shockwaves are not included, since the equations are not based on the fundamental diagram.

2.8 CONCLUSION

At the start of the literature study, a set of questions were presented about the aspects that should be considered when constructing a new control method. Previous studies have given more information about these aspects and are used to answer the research questions. This leads to requirements and wishes for the control method which are mentioned next.

2.8.1 ANSWERING OF QUESTIONS

Which network effects can be identified, when do they occur and how large are the effects? The literature distinguishes the network effects capacity drop, shockwaves, blocking back effects and choice behaviour. The first three effects influence the flow directly, while the route choice behaviour is based on the situation in the network due to these effects of congestion. The capacity drop, shockwaves and blocking back reduce the capacity of the network with 6 to 30%; the effects of choice behaviour are not quantified in literature.

Which DTM measures can be identified and what can be the effects on the network performance? The mostly used DTM measures are dynamic route guidance (DRG), ramp metering and variable speed limits (VSL). These measures are often applied in modelling studies and in practice where DRG and ramp metering give good results, but the effectiveness of VSL is proven less. Another measure mentioned in literature is lane control. Lane control is mostly applied in the form of peak lanes which gives a great improvement: it decreases the congestion weight with 50%, but is only possible when space and budget is available.

Which features are of importance to reach the optimum? Feedback control can give a higher network performance than feed forward control, what means that using feedback control systems can contribute to reach the optimum. Coordination and integration of DTM measures give an additional improvement of the traffic performance with average 5 to 10% compared with local control.

What are the strengths and weaknesses of the current control approaches? For freeway networks four different control approaches are identified: optimal control, model predictive control (MPC), rule-based control and case-based control. Optimal control gives a numerical optimal solution, but does not use feedback and can only be implemented real-time when an efficient algorithm is constructed. Model predictive control has all the advantages of the optimal control and does use feedback but needs much computation time.
what makes it not suitable for large networks or an hierarchical approach. This disadvantage is solved for urban networks by simplification of the network and recasting the problem as a MILP problem. This reduces the computation time by 100 to 400 times, but is not applied yet at large freeway networks.

Rule-based control and case-based control both give fast approximations of optimal control, but will never determine the optimal situation is the network.

**What are the strengths and weaknesses of a static versus dynamic approach?**

A static approach of the traffic situation is less complex than the dynamic approach, but does not represent important characteristics of the traffic network such as a varying demand, blocking back effects and route choice. The static approach can be used for theoretical studies, but is not suitable for (the modelling of) real networks.

**Which traffic models can be used to describe large freeway networks?**

A distinction can be made between first- and higher-order models and here is important to note that more network details in the model lead increase the mathematical complexity and the needed computation time. First order models are the LWR model, the Cell Transmission Model (CTM), the Link Transmission Model (LTM) and the Piecewise Affine Network (PAN) model. The first three models are based on the kinematic wave theory and include shockwaves. The LWR model has some disadvantages due to the simplicity: unrealistic queuing and driving behaviour. The CTM and LTM do represent blocking back effects, but also do not include the capacity drop. Compared to the CTM, LTM offers a less detailed model that is faster but still serves the same traffic features. The PAN model is a static model which is not based on the fundamental diagram, but is designed for control methods. It is the simplest model of all, where the capacity drop is included but shockwaves and blocking back effects are not taken into account. The higher-order models of Payne and Helbing will be too complex for this research since this will result in high computation times, which means that real time control is not feasible.

### 2.8.2 Required and desired properties of control method

Based on the literature, the requirements and wishes for the control method can be specified. Figure 2.11 identifies all the requirements (red boxes) and wishes (black boxes) for the control method, which will be explained further in this section.

![Figure 2.11: Overview requirements control method](image)
Research has shown that the network effects capacity drop, shockwaves and blocking back can all be quantified but the effects of choice behaviour on the traffic state have not been proven yet. Because of this, and because representing choice behaviour will increase the complexity of the traffic model, the last effect will be excluded from the control method. It is desired that all the other three effects will be included in the traffic model, but this depends on the traffic model what makes it “wishes” and not requirements. The DTM measures dynamic route guidance, ramp metering and variable speed limit are proven to be effective and easy to implement. Exactly which measures will be used to control the traffic will depend on the way in which they can be implemented to the control method. To approximate the optimal solution best with the control method, to and improve the network performance as much as possible, it is discussed that coordinated and integrated control is required. Also the use of feedback can improve the solution, which will therefore also be a requirement for the control method. The last requirement is the use of a dynamic approach, because there must be dealt with the changing situation in a traffic network when the control method is used for real time control.
3. Methodology

The literature study has specified the requirements and wishes for the control method, and has researched the different types of control methods and traffic models. In this chapter, the requirements and wishes are translated into the control approach and the traffic model that is suitable for this control approach and meets the requirements and wishes of the previous chapter.

At first the control approach will be presented for a real time control method for large networks, and the requirements for the traffic model will be formulated. Based on these demands, and the traffic models as described in the previous chapter, the modelling approach is presented in the second paragraph. Since the method must be verified and validated, the way in which this will be done is described in the third section. At last, an overview of the methodology is given.

3.1 Control Approach

The objective of the study is to develop a method which can control large networks real time, which gives the interpretation of the first two choices. The literature study has discussed a few real time control methods for large networks, such as the AMOC algorithm (Kotsialos & Papageorgiou, 2001). Since these algorithms do not use feedback, they do not meet the requirements for this research. Further, the control method must enable coordination and integration of the controllers.

The literature study has shown that a closed-loop control approach for freeway networks is model predictive control (MPC). But traditional MPC cannot deal with the real time control of large networks due to the high computation complexity. This disadvantage can be reduced by recasting the problem as a MILP problem have to be used to deal with all the requirements. This approach is applied before in urban networks in the studies of van den Berg et al. (2008) and Lin (2011).

There focus must be on two ways to decrease the needed computation time (Lin, 2011):

1. Reduction of the model
2. Reformulation of the optimisation problem

![Diagram](image)

**Figure 3.1: Set-up control method**

The first action is mentioned earlier: the higher the complexity of the traffic model, the longer the computation time of the optimisation will be. From this point of view, the simplest, yet sufficient, model approach must be chosen to describe the state in the traffic network.

In the second step, the initial non-linear and non-convex optimisation problem for the MPC controllers in the traffic network is reformulated as a mixed integer linear programming (MILP) problem. Solving a MILP problem can, in some cases, require less computation time than non-linear,
non-convex problems what can make the optimisation suitable for real-time application. Both approaches are NP hard but when the optimisation problem can be translated into an efficient MILP problem, the computation complexity can be reduced. To guarantee this reformulation, it is required that the traffic model has a linear or piecewise affine cost function, which sets an extra requirement to the traffic model. The relations between the simplified model and the optimisation are shown in figure 3.1.

This control approach is used in previous studies with the S* model for urban networks (Lin, 2011), but since a freeway network adopts more effects the reformulation of the problem is assumed to be more complex. The reformulation of the problem in freeway networks will be based on these previous studies, where attention must be paid to the existence of the additional effects such as shockwaves and the capacity drop.

3.2 Modelling approach

The control approach as developed in the previous paragraph needs a simple traffic model that minimizes the computation time and has a piecewise affine cost function. Further, there is stated that the model must be dynamic and able to implement the DTM measures ramp metering and route guidance. It is desired that the model represents all the network effects as identified in paragraph 2.2, but should at least consider the capacity drop, shockwaves and blocking back effects.

The theoretical models that are discussed in paragraph 2.6 are evaluated on these requirements and desires; an overview is given in table 3.1. This shows that only the Payne and Gas-Kinetic model meet all the network effects, but the complexity is too high for the control method. The link transmission model (LTM) is considered to be the best overall option but can only be used when it is, or can be made, piecewise affine. Appendix B shows detailed exploration of the LTM, on which is based that the model is not suitable for this research due to the high complexity.

<table>
<thead>
<tr>
<th>Model</th>
<th>Network effects*</th>
<th>Static/dynamic</th>
<th>Complexity**</th>
<th>Simulation software</th>
</tr>
</thead>
<tbody>
<tr>
<td>LWR</td>
<td>-</td>
<td>-</td>
<td>Dynamic</td>
<td>2</td>
</tr>
<tr>
<td>CTM</td>
<td>-</td>
<td>+</td>
<td>Dynamic</td>
<td>4</td>
</tr>
<tr>
<td>LTM</td>
<td>-</td>
<td>+</td>
<td>Dynamic</td>
<td>3</td>
</tr>
<tr>
<td>Payne</td>
<td>+</td>
<td>+</td>
<td>Dynamic</td>
<td>5</td>
</tr>
<tr>
<td>Gas-Kinetic</td>
<td>+</td>
<td>+</td>
<td>Dynamic</td>
<td>6</td>
</tr>
<tr>
<td>PAN model</td>
<td>+</td>
<td>-</td>
<td>Static</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.1: Overview theoretical models
* CD= Capacity drop, SW= Shockwaves, BB= Blocking back, CB= Choice behaviour
** The degree of complexity: 1 is least complex, 6 is most complex

This means that a simpler model must be chosen: the LWR model or the PAN model. Since the PAN model is already a piecewise affine model, it has a big advantage over the LWR model, but also shortcomings such as the fact is it a static model and the network effects shockwaves and blocking back are not included (Hegyi, De Schutter, & Hellendoorn, 2005). Further exploration of this model, as presented in chapter 4, shows a way to include spillback and dynamic demands what makes the model suitable for the research.
3.3 **Validation set-up**

To answer the last research question “*What is the advantage of the new control method over other control methods in terms of computation time and approximation of the optimal solution?*”, the control method that is designed will be validated in chapter 6. At first, there must be verified that the control method is working properly. This is done analytically by reviewing the output given by the control method: although the exact controller settings and network parameters cannot be calculated by hand, there can be stated if the output is logically valid. To ensure the robustness of this verification, several cases will be optimised by the new control method and reviewed.

It is also of importance that the new method will contribute to the current control methods. There is stated that recasting the optimisation problem as a MILP problem can compute the controller settings much faster than the traditional MPC controller and that this should allow the real time optimisation of bigger networks. The extent to which the MLP approach can improve the computation time needed, and the consequences this will have for the maximum network size must be researched to determine the value of the new method compared with the current control methods. This will be done in the second part of the validation.

3.4 **Overview**

Figure 3.2 gives a total overview of all the subjects that are discussed in this chapter.

Since the control method requires a traffic model, this will be constructed first in the next chapter. Here also choices are be made about the network effects that are represented and the DTM measures that are used to control this network.

Next the efficient MPC for this model and the optimisation problem are constructed according to the control approach as described in this chapter. At last, the method will be validated.

---

**Figure 3.2: Schematized overview methodology**

Large networks  Coordination and integration  Online control  Feedback control

Capacity drop  Shortwaves  Spillback

DTM measures  Dynamic approach

MPC via MILP

Extended PAN model

Developing control method

Validation
4. PIECEWISE AFFINE NETWORK MODEL

The PAN model as described by (Hegyi, De Schutter, & Hellendoorn, 2005) adopts control measures in a freeway network model that takes bottlenecks with a capacity drop into account. This model is coupled to a control approach to find ‘the optimal control inputs that minimize the total time that vehicles spend on a freeway network with bottlenecks and capacity drop behaviour’.

The basic PAN model as described in the article of Hegyi et al. (2005) is presented in the first section, followed by the required extensions to make the model suitable for this research.

4.1 ORIGINAL MODEL

We represent the evolution of the traffic network over time in stages where the index \( k \) refers to the \( k \) -th stage. The time duration of stage \( k \) is denoted by \( t(k) \) where \( t(k) \geq 0 \) and fixed.

The benefit of having \( t(k) \) as an independent variable is that - as we will see later - for a given \( k \) the system behaves linearly as a function of the control inputs \( u(k) \) and \( t(k) \). This results in a simpler description of the dynamic behaviour of the system.

However, in practice most traffic control devices, such as ramp metering or route guidance, accept inputs at discrete time steps, e.g., every 30 or 60 seconds. The expectation is that the optimal switching sequence (the moment that the controller settings will change) will result in time durations \( t(k) \) that are significantly larger than the sampling time of the control measures. Therefore the rounding of the time durations \( t(k) \) will result in a small performance loss only.

4.1.1 PROBLEM STATEMENT

The PAN model will be used for the purpose to find the control input sequence that minimises the total time that vehicles spend in the network (TTS). There is chosen for this expression of the network performance because this criterion can be measured in a PWA model.

In order to compare the performance of different control signals we define a time horizon over which the control strategies are compared. Now we are ready to formulate the control problem:

*Given a traffic network as defined, an initial network mode \( M(0) \) and initial queue lengths \( \begin{bmatrix} w^T_e(0) & w^T_b(0) \end{bmatrix} \) at time \( t_0 \) find the control inputs \( u^*_e(k), q^*_n,\text{ext}(k), t^*(k), k = 0, \ldots, K \) that minimize the total time that vehicles spend in the network over \( [t_0, t_0 + t_{hor}] \).*

4.1.2 NETWORK MODEL

The PAN model contains problem specific elements. Each network consists of the following elements (see figure 4.1):

- **origin:**
  \( \bullet \rightarrow q_e \)

- **destination:**
  \( q_d \rightarrow \bullet \)

- **node:**
  \( q_{n,\text{in},1} \leftarrow n \rightarrow q_{n,\text{out},1} \)
  \( q_{n,\text{in},2} \leftarrow n \rightarrow q_{n,\text{out},2} \)

- **bottleneck:**
  \( \bullet \rightarrow q_e,\text{in} \leftarrow \text{in, cap} \rightarrow q_e,\text{out} \)

![Figure 4.1: The network elements (Hegyi, De Schutter, & Hellendoorn, 2005)](image-url)}
- Origin -> Inflow to the network
- Destination -> Outflow of the network
- Node -> Connects one or more incoming links with one or more outgoing links
- Control measure -> DTM measure that limits the outflow
- Bottleneck -> Point where the capacity is limited due to merging, curves etc.
- Links -> Provide the connection between two other elements.

These elements will be explained in more detail in the following sections.

**Origins**

We assume that the origins are the sources of traffic. Origin 0 is an element of the set of all origins \( \{O_1, O_2 \ldots \} \), and provides a constant inflow to the network of \( q_{o}(\text{veh/h}) \).

**Destinations**

Destinations are the sinks of traffic. The average flow at destination \( d \in \{D_1, D_2 \ldots \} \), in stage \( k \), is denoted by \( q_{d}(k) \) (veh/h).

**Nodes**

At nodes traffic from several incoming links may be joined and redistributed over one or two outgoing links. The flows of the incoming links of node \( n \in \{N_1, N_2 \ldots \} \), are denoted by \( q_{\text{n,in},i}(k) \) (with \( i \in I_n \), where \( I_n \) denotes the set of indexes of the incoming links of node \( n \)), and the outgoing links are denoted by \( q_{\text{n,out},j}(k) \) (with \( j \in \{1,2\} \) ). The inflows and outflows are related by \( q_{\text{n,out},j}(k) = \beta_{n,j}(k) q_{n}(k) \), where \( q_{n}(k) = \sum_{i \in I_n} q_{\text{n,in},i}(k) \), and \( \beta_{n,j}(k) \) is the fraction of traffic that leaves node \( n \) through link \( j \). Of course, \( \beta_{n}(k) \geq 0 \) and \( \sum_{j \in O_n} \beta_{n,j}(k) = 1 \), where \( O_n \) is the set of indices of leaving links from node \( n \).

If there is no route guidance, a constant turning rate \( \beta_{n,j}(k) = \beta_{n,j} \) is assumed. If there is route guidance at the node, we will consider \( q_{\text{n,ctrl}}(k) = q_{\text{n,out},1}(k) = \beta_{n,1}(k) q_{n}(k) \) as the control variable. There may be bounds on the route guidance signal \( q_{\text{n,ctrl}}(k) \), which are expressed by \( \beta_{n,1,\text{min}} q_{n}(k) \) and \( \beta_{n,1,\text{max}} q_{n}(k) \), with \( 0 \leq \beta_{n,1,\text{min}} \leq \beta_{n,1,\text{max}} \leq 1 \).

This leads to the following relations:

\[
q_{\text{n,out},1}(k) = q_{\text{n,ctrl}}(k) \tag{4.1}
\]
\[
q_{\text{n,out},2}(k) = q_{\text{n,in}}(k) - q_{\text{n,ctrl}}(k) \tag{4.2}
\]
\[
\beta_{n,1,\text{min}} q_{n}(k) \leq q_{\text{n,ctrl}}(k) \leq \beta_{n,1,\text{max}} q_{n}(k) \tag{4.3}
\]

**Flow-limiting control measures**

Traffic control measures, such as ramp metering, main-stream metering, and dynamic speed limits can be represented by a generalized control measure \( c \in \{C_1, C_2 \ldots \} \) that describes the corresponding flow limitation.

A flow limitation can be active or inactive. The flow limitation is called active when the outflow is limited by the controller (see also figure 4.2 left). So, when it is active there will be a queue present of length \( w_{c}(k) \). The following relations hold:

\[
u_{c}(k) = q_{\text{c,out}}(k) \tau(k) \tag{4.4}
\]
\[
u_{c}(k) \leq q_{\text{c,in}}(k) \tau(k) + w_{c}(k) \tag{4.5}
\]
\[
u_{c}(k) \geq 0 \tag{4.6}
\]
\[
w_{c}(k + 1) = w_{c}(k) + q_{\text{c,in}}(k) \tau(k) - u_{c}(k) \tag{4.7}
\]
where $u_c(k)$ is the control input and represents the number of vehicles that leave the control measure in stage $k$, $q_{c,\text{in}}(k)$ is the inflow and $q_{c,\text{out}}(k)$ the outflow at measure $c$.

![Diagram of inflow and outflow relations](image)

Figure 4.2: The relations between inflow and outflow of a flow-limiting control measure (left) and of a bottleneck (right)

In addition, the bounds on the control input are defined as:

$q_{c,\text{min}}(k) t(k) \leq u_c(k) \leq q_{c,\text{max}}(k) t(k)$ and $0 \leq q_{c,\text{min}} \leq q_{c,\text{max}}$.

If the flow limitation is not active, then there is no queue and the outflow is not limited:

\[
\begin{align*}
    u_c(k) &= q_{c,\text{in}}(k) t(k) \\
    q_{c,\text{out}}(k) &= q_{c,\text{in}}(k) \\
    w_c(k) &= 0 \\
    w_c(k+1) &= w_c(k)
\end{align*}
\]

4.8 - 4.11

The activity status of $c$ is denoted by $X_c$, which has a value 1 if the control measure is active and 0 if it is inactive.

**Bottlenecks**

A generalized bottleneck $b \in \{B_1, B_2, \ldots\}$ may represent several kinds of bottlenecks, such as on-ramps, bridges, tunnels, curves, grades, shock waves, merges, and bifurcations. The common factor in these bottlenecks is that they have a limited capacity $q_{b,\text{cap}}$, and that there may be a capacity drop if the bottleneck is jammed. The queue discharge rate is denoted by $q_{b,\text{dch}}(\leq q_{b,\text{cap}})$, where equality holds if there is no capacity drop, but only a limited capacity.

Similarly to flow-limiting control measures, a bottleneck can also be active or inactive, and the relation between the inflow and outflow depends on the activity status. The basic idea for the bottleneck modelling is that if the inflow exceeds the capacity then the bottleneck will become active (congested) and the outflow will drop to the queue discharge rate (see also figure 4.2 right).

In order to resolve the jam at the bottleneck the inflow must be limited to a value lower than the outflow (the queue discharge rate) and the queue length must be zero. When the jam is resolved, the bottleneck becomes inactive and the outflow may increase up to the capacity again.

So, if the bottleneck is active, then:

\[
\begin{align*}
    q_{b,\text{out}}(k) &= q_{b,\text{dch}} \\
    w_b(k) - (q_{b,\text{dch}} - q_{b,\text{in}}(k)) t(k) &\geq 0 \\
    w_b(k+1) &= w_b(k) - (q_{b,\text{dch}} - q_{b,\text{in}}(k)) t(k)
\end{align*}
\]

4.12 - 4.14

and if the bottleneck is inactive, then

\[
\begin{align*}
    q_{b,\text{out}}(k) &= q_{b,\text{in}}(k) \\
    q_{b,\text{in}}(k) &\leq q_{b,\text{cap}}(k) \\
    w_b(k) &= 0 \\
    w_b(k+1) &= w_b(k)
\end{align*}
\]

4.15 - 4.18
The activity status of \( b \) is denoted by \( X_b \), which has a value 1 if the bottleneck is active and 0 if it is inactive.

**Links**

Links provide the connection between any two other elements. A link connects the outflow of the upstream element with the inflow of the downstream element. The capacity of a link is assumed to be unlimited. If a freeway link with limited capacity is modelled, a bottleneck element should be included. Now we can build networks with the elements from this section.

### 4.1.3 Network properties

**Network mode**

The network mode \( M \) can be defined as the vector of the activity states of all bottlenecks and control measures:

\[
M = [X_{b1}, X_{b2}, ..., X_{c1}, X_{c2}, ...]^T
\]

The current network mode can be acquired from speed, flow and density measurements at the bottlenecks and control measures. The bottleneck modes cannot be controlled directly, only through the available control measures, which may change the inflow of a bottleneck such that a mode change is triggered.

**Mode changes**

If the network is in a given mode, then other modes may be reached autonomously or by varying the control inputs. A queue of a control measure or bottleneck may become zero, which may cause a state change from active to inactive, or if the control inputs are changed, then the activity state of a control measure or a bottleneck may change.

An activity state change is triggered if the inflow of a measure or bottleneck violates the condition that would guarantee the current state (e.g., the condition that for an active bottleneck the inflow must be higher than the queue discharge rate) and if the corresponding queue is zero.

In addition to the relations of the previous section for the (current and next) modes at stages \( k \) and \( k + 1 \), the following relations must hold for the mode change.

If a bottleneck changes from active to inactive then:

\[
\begin{align*}
\omega_b(k) - (q_{b,\text{dch}} - q_{b,\text{in}}(k))t(k) &= 0 \\
q_{b,\text{in}}(k + 1) &\leq q_{b,\text{dch}}
\end{align*}
\]

and if the activity state changes from inactive to active then:

\[
q_{b,\text{in}}(k + 1) > q_{b,\text{cap}}
\]

If a control measure changes from active to inactive then:

\[
\omega_c(k) - u_c(k) + q_{c,\text{in}}(k)t(k) = 0
\]

and there are no additional constraints for control measures if the activity state changes from inactive to active.

Note that it is necessary to require strict inequalities for the bottleneck activity changes since the transition will be triggered only in that case. Since in practice the state transition of bottlenecks (when congestion is created or resolved) may be influenced by disturbances and stochastic effects, we include an extra margin \( _\_ \) in the inequalities related to the mode changes. E.g., the inequality of the change from inactive to active becomes:
For the same reason we also include a constraint to require that each mode has a minimum time length:

\[ t(k) \geq t_{\text{min}} \]

### 4.2 Extended Model

Since the PAN model is a static model which does not represent blocking back effects and shockwaves, the model must be extended to meet all the requirements as stated in the literature study. When doing so, it is important to keep the goal of the research in mind. This means that the extended version of the model must be (kept) piecewise affine, in order to remove all the nonlinearities from the model in the next stage, and as simple as possible.

The model will only be extended with blocking back effects and not shockwaves, because of the complexity to represent shockwaves in a piecewise affine model. Next, this version of the PAN model will only use the flow limiting measures ramp metering and mainstream control. The possibilities to expand the model with dynamic route guidance will be mentioned briefly in the last chapter of this report.

In this paragraph, the extension of the static PAN model with time-varying demand and blocking back effects will be discussed in the first two sections. This is followed by, the way in which the extended PAN model deals with the controllers. The last section gives all the model equations of the extended PAN model.

#### 4.2.1 Piecewise Constant Demand

The paper of van den Berg et al (2009) describes the extension of a static route choice model to the case of time-varying demand profiles. The demand is a piecewise constant function (figure 4.3) during a time period and gives piece wise affine functions for the average time vehicles spend in the network. This information will be used for the extension of the PAN model.

![Figure 4.3: Piecewise constant demand function](image)

The piecewise constant demand function as pictured in figure 4.3 uses the same constant time periods as the original PAN model, displayed as \( t_k \). The main characteristic of a PWA model is the fact that the stage of a link (congested/ free flow) or bottleneck (active/ inactive) cannot change during period \( k \). This means that for every \( k \), there must be determined if a link is congested or not, and if a bottleneck is active or inactive.
The piecewise constant demand pattern that determines the inflows into a network also flows into the links downstream, but with a delay and the flow can change. This is illustrated in example 4.1.

**Example 4.1**

Suppose a simple network with 3 links and a bottleneck in between link 1 and 2 (figure 4.4).

![Network diagram](image)

**Figure 4.4: Lay out network**

In this example, the following values are considered:

Length link 1: \( L_1 = 2 \) km;
Length link 2: \( L_2 = 2 \) km;
Length link 3: \( L_3 = 2 \) km

Free flow speed: \( v_{ff} = 120 \) km/h

Capacity bottleneck: \( q_{cap,b} = 1800 \) veh/h (no capacity drop)

The demand function is piecewise constant and changes every 5 minutes, where:

- \( D_1 = 1800 \) veh/h
- \( D_2 = 2400 \) veh/h
- \( D_3 = 1200 \) veh/h

It is assumed that link 1 is in free flow conditions at the start, what gives a delay of \( \frac{2}{120} = \frac{1}{60} \) h (or 1 minute) between link 1 and link 2. The piecewise constant demand equals the inflow of link 1 as shown in figure 4.5.

![Demand/inflow link 1](image)

**Figure 4.5: Demand/inflow link 1**

The demand at the second link is the same pattern as the demand of link 1, but includes a delay of \( \frac{2}{120} = \frac{1}{60} \) hour (or 1 minute) as shown in figure 4.6. Since the bottleneck can only discharge 1800 vehicles per hour, the actual inflow into link 2 is limited during the 2nd period. Figure 4.6 shows that it takes more than 5 minutes for link 2 to process the demand during this period, and the demand of the last period cannot flow into the link until all the vehicles of the previous period are left.
4.2.2 Blocking back effects

Blocking back effects can be added to the model by setting constraints to the number of vehicles that can leave a link during a time period: no more vehicles can leave a link than space is left in the link downstream. Three different stages can be identified for each link: free flow, a queue is being built up, or the link in fully congested (the jam density is reached). Blocking back effects occur when a link reaches such a density that the outflow of the link upstream will be limited: when an off-ramp is located here it will be blocked by the spillback of the traffic jam.

The inclusion of blocking back effects into the model must be done in such a way that the model stays piecewise affine, as previous section mentioned the stage of a link cannot change during $t_k$. This means that assumptions must be made about the number of vehicles that can leave a link during period $t_k$ in different stages:

Between link 2 and 3 is no bottleneck located, what means that the vehicles will leave the link $2/120 = 1/60$ hour after they entered. This influences the demand function for the third link: time $T$ where the demand is the same is not always 5 minutes, and the value of the demand also changes as shown in figure 4.7.

This example shows how the piecewise constant demand function is distributed over time over the links downstream. The time horizon will not be included in the PAN model, only time periods will be used to calculate the traffic situation. This means that the moment at which the flow of a specific period enters a link will not be determined by the model, but the example has shown how the moment can be derived when the free flow speed is known. However, when the network is not empty or uncongested at start, this approach to determine the time cannot be used: recommendations for further expansion of the model will be given in the last chapter.
When a link has reached the jam density $\rho_{\text{jam}}$, the inflow $I(k)$ into this link must be equal to or less than the outflow $O(k)$ of this link.

When the sum of the queue $w(k-1)$ on the link at the end of the previous period and the demand from the link upstream $S(k)$ is smaller than the sum of the vehicles that can be on the link $\rho_{\text{jam}}L$ and can leave the link $O(k)$, the link can accept the total demand.

These assumptions are expressed mathematically in equation 4.25 for a fictive network with 3 links and no off-ramps, on-ramps or bottlenecks. This equation shows that the number of vehicles that can enter link 2 during period $t$ depends on the number of vehicles that can flow out of this link during this time period and the number of vehicles that can be stored in queue on link 2. When the number of vehicles that want to leave link 1 ($S_1(k)$) is smaller than the sum of these both factors, they can flow out freely, otherwise a queue will be built up at link 1.

\[
I_2(k) = \begin{cases} 
S_1(k), & \text{if } w_2(k-1) + S_1(k) < \rho_{\text{jam}}L_2 + I_3(k) \\
\min(\rho_{\text{jam}}L_2 + I_3(k) - w_2(k-1), S_1(k)), & \text{if } w_2(k-1) + S_1(k) \geq \rho_{\text{jam}}L_2 + I_3(k) 
\end{cases}
\]

When this situation occurs in a network where an off-ramp is situated at the end of link 1, blocking back effects can occur. This is illustrated with the example below.

**Example 4.2**

A network with a main road consisting of three links (link 1, link 3 and link 4), an off-ramp (link 2) at the end of the first link and a bottleneck between link 3 and 4 is presented in figure 4.8.

![Network layout diagram](image)

**Figure 4.8: Network layout blocking back effects**

Table 4.1 gives the piecewise constant demand pattern for the inflow at origin 1 and 2. The turn rate $\beta$ at the node is 0.2, and the other relevant parameters are given below:

- Bottleneck capacity $q_{\text{cap,B}} = 4000$ veh/h
- Jam density of link 3 $\rho_{\text{jam,3}} = 200$ veh/km
- Link lengths $L_1 = L_2 = L_3 = L_4 = 2$ km

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>0-20</th>
<th>20-40</th>
<th>40-60</th>
<th>60-80</th>
<th>80-100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period k</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Origin 1 (veh/h)</td>
<td>6000</td>
<td>6000</td>
<td>5000</td>
<td>4000</td>
<td>2000</td>
</tr>
</tbody>
</table>

**Table 4.1: PWC demand pattern**

Because the traffic demand at the bottleneck is larger than the capacity during the first two periods, a queue will be built up at link 3. Implementation of these parameters in Matlab shows that at the first period only a queue appears at link 3, but that the queue spills back to the first link at the second period. This is pictured in figure 4.9.
Because of spillback of the queue on link 3, the outflow of link 1 is hindered what results in blockage of the off-ramp: fewer vehicles can enter link 2. Figure 4.10 represents the blocking back effects: the number of vehicles per period that want to use the off-ramp (including the vehicles in the queue on link 1), and the number of vehicles per period that can actually leave the network via the off-ramp.

**Figure 4.10: Relation inflow and demand due to spillback**

### 4.2.3 Controllers

The locations of controllers $q_{c,x}$ are determined in advance (together with the lay-out of the network) and will be implemented in the model at the end of the links. This means that a controller directly influences the outflow of a link: when the controller is active, no more vehicles can leave the link than the outflow of the controller $q_{c,x}(k)t_k$.

### 4.2.4 Model equations

Previous sections have shown that the situation in the network can be calculated based on only 4 varying parameters per link: the queue length $w_x(k)$, the inflow into a link $I_x(k)$, the number of vehicles that want to leave a link $S_x(k)$ and the number of vehicles that can leave the link $O_x(k)$. When a controller is located at the end of a link, a fifth parameter will be added: the controller output $q_{c,x}(k)t_k$. The relations among these parameters and the fixed input (such as the jam density $\rho_{jam,x}$ and the link length $L_x$) for different situations are covered in the equations in this section.

**Queuing**

For every time period $k$, the number of vehicles in queue must be determined. It can easily be stated that the number of vehicles in queue ($w_x$) on a link $x$ at time $k$ is the sum of the number of vehicles in queue at the previous time $(k - 1)$ plus the vehicles that enter the link ($I_x$), minus the vehicles that leave the link ($O_x$): 

$$w_x(k) = w_x(k - 1) + I_x(k) - O_x(k)$$

4.26
FLOW TRANSMISSIONS

The relation between the outflow of one link and the inflow in the link downstream depends on the number of links that are connected at a node. When a node is connecting only two links, the number of vehicles leaving link \( x \) at time \( k \) is the same as the number of vehicles entering link \( x+1 \):

\[
O_x(k) = I_{x+1}(k)
\]

When there is one ingoing link (\( x \)) and two outgoing links (\( x+1 \) and \( x+2 \)):

\[
O_x(k) = I_{x+1}(k) + I_{x+2}(k)
\]

When there are two incoming links (\( x-1 \) and \( x \)) and one outgoing link (\( x+1 \)):

\[
I_{x+1}(k) = O_{x-1}(k) + O_x(k)
\]

When the number of vehicles that can enter link \( x+1 \) is smaller than the sum of the demand of the two incoming links, the outflow of these links will be calculated as expressed in equations below.

\[
O_{x-1}(k) = \frac{S_{x-1}(k)}{S_{x-1}(k) + S_x(k)} I_{x+1}(k)
\]

\[
O_x(k) = \frac{S_x(k)}{S_{x-1}(k) + S_x(k)} I_{x+1}(k)
\]

The equations above are not linear which makes them not suitable for a piecewise affine model. Therefore there is chosen to make the assumption that the demand (sending flow) of one of the two links can always flow freely into the link downstream. For example: when an on-ramp (link \( x \)) is located the sending of this link will flow freely into the main road, while the outflow of the main road upstream (link \( x-1 \)) will be reduced. This is represented by the linear equations below.

\[
O_{x-1}(k) = I_{x+1}(k) - S_x(k)
\]

\[
O_x(k) = S_x(k)
\]

BLOCKING BACK EFFECTS

Equation 4.25 of the previous section determines the effects of spillback on the links upstream, but does not include the capacity drop. Combining these two network effects gives the equation for the inflow into a link:

\[
I_{x+1}(k) = \begin{cases} 
I_{x+1}(k) & \text{if } \left( w_{x+1}(k - 1) + S_x(k) \right) \geq \left( \rho_{jam} I_{x+1} + O_{x+1}(k) \right) \\
\min(\rho_{jam} I_{x+1} + O_{x+1}(k) - w_{x+1}(k - 1), S_x(k)) & \text{if } \left( w_{x+1}(k - 1) + S_x(k) \right) < \left( \rho_{jam} I_{x+1} + O_{x+1}(k) \right) \text{ or } S_x(k) \leq q_{cap,x+1} t_k \\
S_x(k), & \text{or } w_x(k - 1) = 0 \\
\min(q_{disp,x+1} t_k, S_x(k)) & \text{else}
\end{cases}
\]

Here, \( S_x(k) \) reflects the number of vehicles that want to leave link \( x \):

\[
S_x(k) = I_x(k) + w_x(k - 1)
\]
Equation 4.34 only represents a situation with one ingoing and one outgoing link. If an off-ramp is located at the link before a bottleneck, the outflow of this link will also be hindered when the jam spills back (as shown in example 4.2). Equation 4.36 represents the inflow into the on-going link when an off-ramp with fixed turn fraction \(1 - \beta\) is present.

\[
I_x(k) = \begin{cases} 
I_x(k) & \text{if } w_{x+1}(k-1) + \beta S_x(k) \geq \rho_{\text{jam}} L_{x+1} + O_{x+1}(k) \\
\min(\rho_{\text{jam}} L_{x+1} + O_{x+1}(k) - w_{x+1}(k-1), \beta S_x(k)) & \text{if } (w_{x+1}(k-1) + \beta S_x(k)) < (\rho_{\text{jam}} L_x + O_{x+1}(k)) \land S_x(k) \leq q_{\text{cap},x+1} t_k \land w_x(k-1) \\
0 & \text{else}
\end{cases}
\]

4.36

The amount of vehicles that enters the off-ramp \(I_o(k)\) is calculated by equation 4.37:

\[
I_o(k) = \frac{1 - \beta}{\beta} I_x(k)
\]

4.37

When a bottleneck connects two ingoing links and one outgoing link, possible spillback must be addressed to the two links. It will be plausible that the spillback will be divided over the links according to the ratio of the sending flows of these links (when the demand of link 1 is twice the demand of link 2, link 1 will receive 2/3 of the spillback and link 2 only 1/3). The inflow into the link downstream the bottleneck (link3) is captured in equation 4.38.

\[
I_3(k) = \begin{cases} 
I_3(k) & \text{if } (w_3(k-1) + S_1(k) + S_2(k)) \geq (\rho_{\text{jam}} L_3 + O_3(k)) \\
\min(\rho_{\text{jam}} L_3 + O_3(k) - w_3(k-1), S_1(k) + S_2(k)) & \text{if } (w_3(k-1) + S_1(k)S_2(k)) < (\rho_{\text{jam}} L_3 + O_3(k)) \text{ or } S_1(k) + S_2(k) \leq q_{\text{cap},3} t_k \\
0 & \text{or } w_3(k-1) = 0
\end{cases}
\]

4.38

\[
S_1(k) + S_2(k),
\]

\[
\min \left( q_{\text{dch},3} t_k, S_1(k) + S_2(k) \right)
\]

Controlled outflow

When an active controller is located at the end of a link, the outflow of this link can be limited. In this situation, the sending flow of a link must be replaced by the controller outflow \(q_{c,x}(k) t_k\) in the equations 4.30-4.34, 4.36 and 4.38. For the controller outflow, constraints must be set because the outflow can never exceed the sending flow: \(q_{c,x}(k) t_k \leq S_x(k)\).

4.3 Validity check PAN model

The functioning of the extended PAN model is validated in this paragraph. At first the weaknesses will be discussed, followed by a quick scan of the functioning of the model.

4.3.1 Weaknesses of PAN model

Although the model meets all the requirements for this research, it has a few weaknesses that must be mentioned. At first, the model as formulated in this chapter cannot be applied to network with multiple routes. Due to the use of time periods instead of a time horizon, the model cannot handle multiple arrival times at a link when they are not covered in the same PWC period notation.
model equations in this formation use only one time period overall, and when the time period in which vehicles enter from an on-ramp differs, the flow from this link cannot be matched to the flows on the main road. Suggestions how the PAN model can be extended further, to make it suitable for networks with different arrival times, are given in the last chapter.

Another weakness is the fact that the network state cannot change during period $k$: bottlenecks cannot switch from active to inactive (or vice versa). This has consequences for the output of the model; the bottleneck can stay active or inactive in the model while a queue could already be solved or occurred. The model can therefore give higher or lower outflows than in reality. This effect can be reduced by taking small time periods $t_k$: the activity state of the bottleneck can change faster earlier when the time periods are smaller. For example: if a queue can be solved after 5 minutes, but the used time period is 20 minutes, the bottleneck will keep active 15 minutes too long and will unnecessarily reduce the outflow during these 15 minutes. Reducing the length time period to 1 minute means that the bottleneck will actually be deactivated after 5 minutes and the next 15 minutes the full outflow capacity can be used.

At last it must be mentioned that the spillback of congestion does not start before the jam density is reached, which is not consistent with the kinematic wave theory. According this theory, a density over the critical density (at the capacity flow) causes a decreasing flow and speed. This relation is pictured in the fundamental diagram (figure 4.11).

![Figure 4.11: Fundamental diagram](image.png)

This means that the PAN model does not represent the effect of a higher density downstream until that link reached the jam density. Also there is assumed that the outflow of a fully congested link will always equal the discharge capacity of the bottleneck downstream, while the kinematic wave theory would indicate the occurrence of shockwaves in this situation. These assumptions can cause higher outflows of the links in the PAN model compared with models that are based on the kinematic wave theory. However, this does not have to influence the network performance, since the outflow of bottlenecks will be the same.

### 4.3.2 Example

The previous section has mentioned a few drawbacks of the extended PAN model. It is of importance to check if the overall performance is still realistic since the control input will be determined by this model.

![Figure 4.12: Lay out simple network for quick validity check](image.png)
Therefore the traffic situation in a simple network (figure 4.12), given by the extended PAN model, will be analysed.

<table>
<thead>
<tr>
<th>Time (h)</th>
<th>0-1/3</th>
<th>1/3-2/3</th>
<th>2/3-1</th>
<th>1-4/3</th>
<th>4/3-5/3</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Demand 1 (veh/h)</td>
<td>5000</td>
<td>5000</td>
<td>4000</td>
<td>3000</td>
<td>2000</td>
</tr>
<tr>
<td>Demand 2 (veh/h)</td>
<td>2000</td>
<td>2000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
</tbody>
</table>

Table 4.2: PWC pattern for the network of figure 4.12

The demand pattern for this network is piecewise constant, and given in table 4.2. The relevant network parameters for this network are given below.

\[
\begin{align*}
q_{\text{cap}} & \quad \text{Bottleneck capacity} & 6000 \text{ veh/h} \\
q_{\text{dch}} & \quad \text{Bottleneck discharge rate (capacity drop)} & 4800 \text{ veh/h} \\
\rho_{\text{jam,2}} & \quad \text{Jam density link 3} & 200 \text{ veh/km} \\
L_3 & \quad \text{Length link 3} & 2 \text{ km} \\
t_K & \quad \text{Time of period } k & 1/3 \text{ h} \\
\beta & \quad \text{Turn fraction (to destination 1)} & 0.1
\end{align*}
\]

Table 4.2 shows that the demand from both origins together will exceed the bottleneck capacity during the first two periods, which will cause a capacity drop. In the same time queues will be built up at both link 3 and link 4, since the inflow into this links are higher than the outflows. The total demand for destination 2 can be calculated by taking the sum of 0.9 times the demand at origin 1 and the demand at origin 2 =8033 vehicles. This is pictured in figure 4.13 by the red line.

![Figure 4.13: Overview flow transmission in PAN model](image)

Figure 4.13 and figure 4.14 show the output given by the extended PAN model. It shows that the bottleneck stays active during the first four periods: the slope represents an outflow of 4800 vehicles per hour. This matches the queue development in the figure below: queues are located at link 3 and link 4 during these periods and will not be solved until the last period is finished. The figure also shows that when the jam density of link 3 is reached, the jam spills back into link 1.

![Figure 4.14: Queue development in PAN model](image)
Both figures indicate that the PAN model gives a logically valid output: queues are built up as long as the demand is higher than the bottleneck capacity and started to resolve when the demand was lower than the bottleneck capacity. Also the blocking back effects are represented: the inflow of link 2 is hindered by the spillback of the jam at link 3. However, the figures also show the effects of the use of time periods on the flows and the queue development: the slopes of the lines do not change during a period. This has consequences for the speed with which a queue is resolved: the PAN model will not solve a queue faster than one period, while this could be possible due to the maximum capacity.

4.4 **Non-linear optimisation problem**

The networks performance in the PAN model can be optimised by several solvers for non-linear programming (NLP) problems, using DTM measures as controllers. The fmincon function in MatLab (also available in other toolboxes) has four algorithm options, of which SQP (sequential quadratic programming) is most suitable for the PAN model due to the linear structure. This algorithm can handle constrained NLP problems and can be applied to the PAN model. All the parameters of the extended PAN model must be covered by the system $x$, which will be specified in the input parts:

- The cost function that needs to be minimised
- Linear constraints and equalities: $Ax \leq b$, $Aeq \cdot x = beq$
- Non-linear constraints and equalities: $c(x) \leq 0$, $ceq(x) = 0$
- Upper and lower boundaries for all the variables in the model: $lb \leq x \leq ub$

This paragraph will translate the optimisation objective of the PAN model into the cost function for the fmincon solver in the first section. This is followed by a description of the other input that is needed. At last there will be explained briefly how fmincon solves this problem.

4.4.1 **Optimisation objective**

The optimisation objective is minimizing the total time spent (TTS) in the network (Hegyi, De Schutter, & Hellendoorn, 2005). This can be estimated by using the total time spent in queue, since minimizing the queues implies the same:

$$J_{TTS}(S,K) = \sum_{k=1}^{K} \frac{w(k) + w(k + 1)}{2} t_k$$

4.4.2 **Constraints, equalities and boundaries**

The equations for the queues $w$, the sending flows $S$ and the outflows $O$ are linear and can be covered by the linear equalities. Vector $beq$ consists of real variables, and $Aeq$ is a matrix that indicates the values of the model parameters belong by the outcome $beq$. When constraints are added to the model, for example a maximum queue length to prevent spillback to the underlying network, which can be covered the same way in the linear constraints.

The non-linear equations that determine the inflow of links must be covered in a separate file in MatLab. This file contains the relevant if/else en min/max functions of the PAN model. Since the PAN model uses only linear inequality constraints, $c(x)$ will be empty.

At last, the vectors $lb$ and $ub$ indicate for each parameters of system $x$ the minimum and maximum values.
4.4.3 SQP Algorithm

SQP is an iterative procedure which models the non-linear programming problem for a given iterate $x^k$, $k \in IN_o$ by a Quadratic Programming (QP) sub problem. It solves that QP sub problem and then uses the solution to construct a new iterate $x^{k+1}$. This construction is done in such a way that the sequence $(x^k)_{k \in IN_o}$ converges to a local minimum $x^*$ of the NLP as $k \to \infty$ (Boggs & Tolle, 1996).

The mathematical complexity of the NLP problems, in terms of number of constraints and number of variables, influences the number of iterations are needed to find the local minimum: when the complexity increases, the number of iterations increases which will result in larger computation times.

4.5 Overview

The methodology for the modelling approach indicates a few requirements for the traffic model that can be used for the control method: it must handle a dynamic traffic demand and it must be able to represent blocking back effects and the capacity drop. Because the original PAN model uses static demands and does not include realistic queuing, the model is extended in this chapter. This has resulted in a generalised store-forward model that does meet all the requirements, including the simplicity that is needed for the transformation in the next chapter.

The strength of the model lies in its simplicity: this is required for the control method, but the model also has a few weaknesses. The most important weaknesses are the consequences of the use of time periods and fact that the traffic situation downstream does not influence the links upstream until the jam density of links is used. This means that the extended PAN model as formulated in this chapter cannot be used for networks with multiple routes, and the output of the model is less realistic. The last fact can be compensated by the use of small time periods. However, there is shown that the overall output of the PAN model seems logically valid for the case of section 4.3.2.
5. CONTROL METHOD

The complete control method that uses mixed integer linear programming to solve the optimisation problem of the PAN model is pictured in figure 5.1. The PAN model is constructed in the previous chapter and the next steps will be performed in this chapter.

At first, the PAN model must be transformed into a linear, constrained model to suit the MILP optimisation. This is done by the use of logic rules, which will be mentioned and explained in the first paragraph of this chapter. These rules are used to transform the extended PAN model: the principles of the transformation is given in the second paragraph, the complete transformation is given in appendix C. After the first two paragraphs the linear, constrained PAN is formulated and there will be focussed on the MILP optimisation. The third paragraph describes how a MILP problem can be constructed, which program will be used to solve this problem, and how the belonging algorithm functions.

![Figure 5.1: Schematization of the total control loop](image)

These three paragraphs together produce the bottom part of figure 5.1, and can be applied as MPC controller to another (more extensive) traffic model or even to reality. This step is not developed in this research, but similar studies such as (Lin, 2011) and (van den Berg, De Schutter, Hegyi, & Hellendoorn, 2008) show that it is possible to use the bottom part as MPC controller.

5.1 RULES FOR TRANSFORMATION

The extended PAN model can be transformed into a linear model by the use of logical rules. This paragraph first gives some background theory, followed by a description how non-linear equations can be made linear by the use of these logical rules.

5.1.1 PRELIMINARIES

Some basic tools are introduced for transforming logical statements involving continuous variables into mixed-integer linear inequalities (Bemporad & Morari, 1999).

Capital letters $X_i$ are used to represent statements, e.g. “$x \leq 0$” or “Color is black”. $X_i$ is commonly referred to as a literal, and has a truth value of either “T” (true) or “F” (false). Boolean algebra enables statements to be combined in compound statements by means of connectives:

- “$\wedge$” — and;
- “$\lor$” — or;
- “$\neg$” — not;
- “$\Rightarrow$” — implies;
- “$\Leftrightarrow$” — if and only if;
- “$\oplus$” — exclusive or.
Table 5.1: Truth table (T= true, F=false)

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_1 \land X_2$</th>
<th>$X_1 \lor X_2$</th>
<th>$\neg X_1$</th>
<th>$X_1 \Rightarrow X_2$</th>
<th>$X_1 \Leftarrow X_2$</th>
<th>$X_1 \Leftrightarrow X_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
</tbody>
</table>

These connectives are defined by means of the truth table given in table 5.1. The following prosperities will be used later on:

- $X_1 \Rightarrow X_2$ is the same as $\neg X_1 \lor X_2$  
- $X_1 \Rightarrow X_2$ is the same as $\neg X_2 \Rightarrow \neg X_1$  
- $X_1 \Leftrightarrow X_2$ is the same as $(X_1 \Rightarrow X_2) \land (X_2 \Rightarrow X_1)$

One can associate with a literal $X_i$ a logical variable $\delta_i \in \{0,1\}$, which has a value of either 1 if $X_i = T$, or 0 if $X_i = F$. The following propositions and linear constraints are then equivalent:

- $X_1 \land X_2$ is equivalent to $\delta_1 = \delta_2 = 1$  
- $X_1 \lor X_2$ is equivalent to $1 \leq \delta_1 + \delta_2 \leq 2$  
- $\neg X_1$ is equivalent to $\delta_1 = 0$  
- $X_1 \Rightarrow X_2$ is equivalent to $\delta_1 - \delta_2 \leq 0$  
- $X_1 \Leftarrow X_2$ is equivalent to $\delta_1 - \delta_2 = 0$  
- $X_1 \Leftrightarrow X_2$ is equivalent to $\delta_1 + \delta_2 = 1$

This computational inference technique can be used to model logical parts of processes (on/off switches, discrete mechanisms, combinational and sequential networks) and heuristics knowledge about plant operation as integer linear inequalities.

### 5.1.2 Examples of Transformation

These preliminaries can be used to transform a non-linear model into a Mixed Logical Dynamical (MLD) model. This model allows specifying the evolution of continuous variables through linear dynamic equations, of discrete variables through propositional logic statements and automata, and the mutual interaction between the two (Bemporad & Morari, 1999). A MLD can be optimised by a MILP optimisation, and therefore the PAN model must be transformed into a MLD model.

The key idea of transforming a model into an MLD model consists of embedding the logic part in the state equations by transforming Boolean variables into 0/1 integers and by expressing the relations as mixed-integer linear inequalities.

The PAN model contains two types of non-linear equations: min-max functions and if/else statements. The way in which these equations can be transformed into linear one is described below (Bemporad & Morari, 1999).

#### Min-max functions

According to (Christiansen, 1997) consider the statement $f(x) \leq 0$, where $f: \mathbb{R}^n \to \mathbb{R}$. Assume that $x \in X$, where $X \subset \mathbb{R}^n$ is a given bounded set, and define

$$M = \max_{x \in X} f(x) , \ m = \min_{x \in X} f(x)$$

Theoretically, an over-estimate (or under-estimate) of $M$ or $m$ suffices for our purpose. However, more realistic estimates provide computational benefits. Now, by introducing in $\delta \in \{0,1\}$, it is easy to verify that
where $\varepsilon$ is a small tolerance (typically the machine precision), beyond which the constraint is regarded as violated.

The reason for introducing $\varepsilon$ is that equations like $f(x) > 0$ does not fit the mixed integer linear programming framework, in which only unconstric inequalities are allowed. Therefore, the equation $f(x) > 0$ is replaced by the equation $f(x) \geq \varepsilon$ with $\varepsilon$ a small tolerance, typically the machine precision, where we assume that in practice the case $0 < f(x) \leq \varepsilon$ cannot occur due to the finite number of bits used for representing real numbers on a computer. Then, the following equivalence holds:

\[
[f(x) \leq 0] \Rightarrow [\delta = 1] \quad \text{is true if and only if} \quad f(x) - \delta \leq -1 + m(1 - \delta) \tag{5.11}
\]

\[
[f(x) \leq 0] \Leftrightarrow [\delta = 1] \quad \text{is true if and only if} \quad f(x) \leq M\delta \tag{5.12}
\]

\[
\sim (f(x) \leq 0) \quad \text{is true if and only if} \quad f(x) \geq \varepsilon \tag{5.13}
\]

Moreover, the term $\delta f(x)$, where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and $\delta \in [0,1]$, can be replaced by the auxiliary real variable $z = \delta f(x)$ which satisfies $[\delta = 0] \Rightarrow [z = 0]$, $[\delta = 1] \Rightarrow [z = f(x)]$. Therefore, by defining $M$ and $m$ as in equation 5.10, $z = \delta f(x)$ is equivalent to:

\[
\begin{align*}
z & \leq M\delta \\
z & \geq m\delta \\
z & \leq f(x) - m(1 - \delta) \\
z & \geq f(x) - M(1 - \delta)
\end{align*} \tag{5.14}
\]

\[
[f(x) \leq 0] \Rightarrow [\delta = 1] \quad \text{is true if and only if} \quad f(x) \geq \varepsilon + (m - \varepsilon)\delta \tag{5.15}
\]

Consider the following system:

\[
x(t + 1) = \begin{cases} 
0.8x(t) + u(t) & \text{if } x(t) \geq 0 \\
-0.8x(t) + u(t) & \text{if } x(t) < 0 
\end{cases} \tag{5.17}
\]

where $x(t) \in [-10,10]$, and $u(t) \in [-1,1]$. The condition $x(t) \geq 0$ can be associated to a binary variable $\delta(t)$ such that

\[
\delta(t) = 1 \iff [x(t) \geq 0] \tag{5.18}
\]

By using the transformation rule 5.15, equation 5.18 can be expressed by the inequalities:

\[
\begin{align*}
-m\delta(t) & \leq x(t) - m \\
-(M + \varepsilon)\delta(t) & \leq -x(t) - \varepsilon
\end{align*}
\]

where $M = -m = 10$, and $\varepsilon$ is a small positive scalar.

Then equation 5.17 can be rewritten as

\[
x(t + 1) = 1.6\delta(t)x(t) - 0.8x(t) + u(t) \tag{5.19}
\]
By defining a new variable \( z(t) = \delta(t)x(t) \) which, by equation 5.16, can be expressed as

\[
\begin{align*}
z &\leq M\delta(t) \\
z &\geq m\delta(t) \\
z &\leq x(t) - m(1 - \delta(t)) \\
z &\geq x(t) - M(1 - \delta(t))
\end{align*}
\]

the evolution of system 5.17 is ruled by the linear equation

\[ x(t + 1) = 1.6z(t) - 0.8x(t) + u(t) \]

subject to the linear constraints above.

### 5.2 Transformation PAN model

The equations of the PAN model that calculate the inflow of a link are nonlinear and must be rewritten to perform a MILP optimisation. The equations 4.34, 4.36 and 4.38 can be made linear by the use of logical variables \( \delta \in \{0,1\} \) and auxiliary real variables \( z_i \). The way in which this must be done will be described in this section; the complete reformulation is given in appendix C.

#### 5.2.1 Min-max functions

Equations 4.34, 4.36 and 4.38 all contain a minimum function that must be removed to get a linear expression. Introducing

\[
\begin{align*}
[\delta_1(k) = 1] &\leftrightarrow [\rho_{jam}L_{x+1} + O_{x+1}(k) - w_{x+1}(k - 1) - S_x(k) \geq 0]
\end{align*}
\]

can remove the minimum function of the last two equations:

\[
\begin{align*}
\min(S_x(k), q_{dch,x+1}t_k) &= \delta_1(k)S_x(k) - \delta_1(k)q_{dch,x+1}t_k + q_{dch,x+1}t_k 
\end{align*}
\]

And introducing \([\delta_2(k) = 1] \leftrightarrow [S_x(k) - q_{dch,x+1}t_k \geq 0] \) makes the first equation linear:

\[
\begin{align*}
a(k) &= \delta_2(k) \left( \rho_{jam}L_{x+1} + O_{x+1}(k) - w_{x+1}(k - 1) \right) - \delta_2(k)S_x(k) + S_x(k)
\end{align*}
\]

The constraints for \( \delta_1(k) \) and \( \delta_2(k) \) are given in the appendix.

#### 5.2.2 If/ifelse/else statements

The non-linear equations 4.34, 4.36 and 4.38 all contain an if/ifelse/else statement that must be removed. Based on the transformation rules that describe how to transform a single if/else statement, these equations will be split into two single if/else statements:

\[
\begin{align*}
I_{x+1}(k) &= \begin{cases}
  a(k), & \text{if } (w_{x+1}(k - 1) + S_x(k)) \geq (\rho_{jam}L_{x+1} + O_{x+1}(k)) \\
  b(k), & \text{if } (w_{x+1}(k - 1) + S_x(k)) < (\rho_{jam}L_{x+1} + O_{x+1}(k))
\end{cases}
\end{align*}
\]

Here,

\[
\begin{align*}
a(k) &= \min(\rho_{jam}L_{x+1} + O_{x+1}(k) - w_{x+1}(k - 1), S_x(k)) \\
b(k) &= \begin{cases}
  S_x(k), & \text{if } S_x(k) \leq q_{cap,x+1}t_k \text{ and } w_x(k - 1) = 0 \\
  \min(S_x(k), q_{dch,x+1}t_k), & \text{else}
\end{cases}
\end{align*}
\]

Because equation 5.24 still contains two conditions, this must also be split up:

\[
\begin{align*}
b(k) &= \begin{cases}
  c(k), & \text{if } S_x(k) \leq q_{cap,x+1}t_k \\
  \min(S_x(k), q_{dch,x+1}t_k), & \text{if } S_x(k) > q_{cap,x+1}t_k
\end{cases}
\end{align*}
\]

\[
\begin{align*}
c(k) &= \begin{cases}
  S_x(k), & \text{if } w_x(k - 1) = 0 \\
  \min(S_x(k), q_{dch,x+1}t_k), & \text{if } w_x(k - 1) > 0
\end{cases}
\end{align*}
\]

Equation 5.22 can be rewritten using \([\delta_1(k) = 1] \leftrightarrow [\rho_{jam}L_{x+1} - w_{x+1}(k - 1) \geq 0] \):
\[ I_{x+1}(k) = \delta_2(k) b - \delta_1(k) a + a \]  

This can also be done for \( b(k) \) and \( c(k) \) using \( [\delta_2(k) = 1] \leftrightarrow [q_{cap,x+1} t_k - S_x(k) \geq 0] \) and \( [\delta_3(k) = 1] \leftrightarrow [w_x(k-1) = 0] \):

\[
\begin{align*}
b(k) &= \delta_2(k) c - \delta_2(k)(\min(S_x(k), q_{dch,x+1} t_k)) + \min(S_x(k), q_{dch,x+1} t_k) \\
c(k) &= \delta_3(k) S_x(k) - \delta_3(k)(\min(S_x(k), q_{dch,x+1} t_k)) + \min(S_x(k), q_{dch,x+1} t_k)
\end{align*}
\]

The constraints belonging to the binary variables \( \delta_1(k), \delta_2(k) \text{ and } \delta_3(k) \) are given in appendix C.

5.2.3 Overview

Combining all the reformulations of the previous sections gives an extensive equation for the inflow (given in the appendix), which still contains non-linear parts such as \( \delta_1(k) a \text{ and } \delta_1(k) \delta_2(k) \). For the first type, auxiliary variables \( z_1(k) \) will be used to replace the non-linear expressions, using a set of constraints to ensure the correct value. The second type can be replaced by a new binary variable \( \delta_1(k) \), which can only be 1 when all the binary variables that are kept in the expression have value 1.

The complete reformulation with all the constraints for the binary and auxiliary variables can be found in the appendix and leads to the final linear expression for the inflow:

\[
I_{x+1}(k) = z_1(k) - z_2(k) + \delta_3(k) q_{dch,x+1} t_k - \delta_3(k) q_{dch,x+1} t_k + z_3(k) - \delta_3(k) q_{dch,x+1} t_k + \delta_3(k) q_{dch,x+1} t_k - \delta_3(k) q_{dch,x+1} t_k + z_3(k) + z_4(k) - z_5(k) + \delta_1(k) p_{lam} l_{x+1} + z_6(k) - z_7(k) - z_8(k) + S_x(k)
\]

5.3 MILP optimisation

The previous paragraph has shown how the extended PAN model can be transformed into a MLD model. The settings for the DTM measures in this model can be determined by formulating the model as a mixed integer linear programming problem, given an optimisation objective and a set of constraints.

This paragraph will describe how the MILP problem must be constructed, which program is used to solve the problem and how the solver functions mathematically.

5.3.1 MILP input

To solve the controlling problem as a MILP problem, the MLD model must be formulated as a set of vectors and matrices, where system \( x \) represents the total of all the real, binary and auxiliary variables of the PAN model:

- Cost function which must be minimized: \( f \) (\( nx \) by 1 matrix)
- Equalities: \( Aeq \times x = beq \) (\( Aeq \) is \( ne \) by \( nx \) matrix, \( beq \) is \( ne \) by 1 matrix)
- Constraints: \( Ax \leq b \) (\( A \) is \( nc \) by \( nx \) matrix, \( b \) is \( nc \) by 1 matrix)
- Lower bound (\( lb \)) and upper bound (\( ub \)) of system \( x \) (both \( nx \) by 1 matrix)
- Definition which variables are integers: \( intvars \) (integer vector)

The cost function, which represents the optimisation objective, in a MILP problem must be linear. Since the original cost function, as formulated in chapter 4, is already linear this can be kept:

\[ J_{TTS}(S,K) = \sum_{k=1}^{K} \frac{w(k) + w(k+1)}{2} t_k \]
The set of equalities represent all the equations of the extended PAN model, including the equations that are transformed in the previous paragraph. The constraints belonging to the transformed equations are covered in the set of constraints. The control constraints, such as the maximum queue length or the minimal outflow of a link, are not included in $Ax \leq b$ but in the lower and upper boundaries of the system. Since flows of queues can never be negative, the lower boundary of all the variables must be 0 (unless determines otherwise, such as the minimum outflow). The upper boundaries of the real and auxiliary variable are not always known, in that case a large number must be chosen. For all the binary variables, the lower bound is set on 0 and the upper bound on 1. In the vector intvars these variables are indicated by a ‘1’ (otherwise ‘0’), which means that the variables can only be an integer number. Together with the boundaries, this gives the binary variables which can only be 0 or 1.

When all these vectors and matrices are constructed, the MILP problem can be implemented into a numerical computing environment such as Matlab or Java. The programs and toolboxes that are used to solve this MILP problem will be mentioned in the next section, as well as the way in which the problem is solved mathematically.

### 5.3.2 Solving the MILP Problem

The original PAN model as well as the transformed, MLD model are implemented in Matlab. Since Matlab does not have a solver for MILP problems, the cplex toolbox of TOMLAB is used to solve the control problem. This toolbox uses the Matlab interface and can solve linear and quadratic programming problems as well as mixed integer linear and mixed integer quadratic programming problems. When all the vectors and matrices, as defines in the section above, are implemented in Matlab the cplex solver can be called. This will give a set of output variables and vectors, of which the following are most relevant to mention here:

- All values of the parameters of system $x$
- Value of the cost function: $f_{.k}$
- The result of the cplex run: *Inform*
  - The solution is found
  - The solution is integer infeasible
  - Time limit exceeded
  - Etc.

The cplex toolbox uses the branch-and-bound algorithm to solve the MILP problem. The branch-and-bound algorithm solves a mixed integer linear program by dividing the search space and generating a sequence of sub problems. The steps of this algorithm are discussed briefly below (Dua & Pistikopoulos, 2000):

1. The linear relaxation LP problem of the original MILP problem is solved by the dual simplex algorithm, which gives an optimal solution
2. There is checked if the solution is integer feasible: when it is feasible, the MILP solution is found, otherwise step 3 follows.
3. Sub problems are constructed for each integer wherefore the LP solution is not feasible. Now the search space of a mixed integer linear program can be represented by a tree. Each node in the tree is identified with a sub problem derived from previous sub problems on the path leading to the root of the tree. This is called branching.
4. In this step the branch-and-bound algorithm chooses one of the active nodes and attempts to solve the linear programming relaxation of that sub problem. If the sub problem can be
solved and the solution is integer feasible, then its objective value provides an upper bound for the objective value in the minimization problem (MILP). If the solution is not integer feasible, then it defines two new sub problems and this step will be performed again.

5. Branching continues in this manner until there are no sub problems left. At this point the best integer solution found is an optimal solution for (MILP). If no integer solution has been found, then (MILP) is integer infeasible.

This algorithm shows that the computation complexity of a MILP problem depends first on the number of constraints, as this is important for the dual simplex algorithm in step 1. Next, the complexity depends on the number of integers, since each integer can result in two sub problems that must be solved.

5.4 Overview

This chapter described how the extended PAN model can be made suitable for a MILP optimisation and how this MILP problem can be solved.

By the use of logic rules every min/max function and every if/else statement can be transformed into a linear, constrained equation. The transformation is performed by hand in this research but can also be programmed in MATLAB (or other computation environment), which means that the program can determine the input for the MILP problem based on the network lay-out. This will reduce the human time needed for the use of this control method.

The MLD model that is constructed in the second paragraph can be implemented in MatLab to solve it as a MILP problem. Since MatLab cannot solve MILP problems there is chosen to use the TOMLAB cplex toolbox, which uses the branch-and-bound algorithm to solve the problem. The computation complexity of this algorithm increases when the MILP problem contains more integers, since each integer can result in two new sub problems that must be solved by the algorithm.

At last there must be recalled that this chapter only gives the lower part of the control scheme of figure 5.1. This means that the control method is not useful yet as MPC controller to apply to another traffic network or reality. This will be discussed further in the next chapter and will give recommendations for further expansion of the control network in the last chapter.
6. Verification of the approach and projection of the performance for larger networks

The control method as developed in the previous chapters must be validated according to the last research question: “What is the advantage of the new control method over other control methods in terms of computation time and approximation of the optimal solution?”

To answer this question the requirements for the control methods must be recalled first, as well as the desired application of the method. The methodology for the control approach (chapter 3) identifies the following requirements for the control method:

1. Representation of the capacity drop and blocking back effects.
2. Approximate the optimal solution as much as possible by:
   a. Use of coordinated and integrated control
   b. Use of feedback control
3. Suitable for the real time control of large traffic networks

Whether the control method meets the first two requirements can be checked easily, by analysing the functioning of the control method when it is used for coordinated and integrated feedback control. This means that the solution of the MILP optimisation must be logically valid for a situation with multiple controllers of different types and the MILP optimisation should be used as MPC controller. The functioning of the control method will therefore be verified in the first paragraph of this chapter. Applying the MILP optimisation as MPC controller is outside the scope of this research, but the studies of Lin (2011) and van den Berg et al (2008) have shown that (although it is time consuming) it is possible to couple the MILP optimisation as MPC controller to another model. Based on these studies, there is assumed that it will also be possible for this control method.

The third requirement focuses on the relationship between the computation time needed for the optimisation and the size of the traffic network. Literature has shown that, in general, increasing the network size leads to longer computation time. Since the computation time of the optimisation must be smaller or equal to real time to make the control method suitable for real time control, there will be a maximum network size that can be controlled by the method. The dimension of the network (corridor, city, county) to which the control method can be applied is of importance to determine the value of this control method over others. Therefore the relation between model size and computation time will be researched in the second paragraph.

At last, the results of the verification and the study after the computation complexity will be combined to answer the research question as stated above.

6.1 Verification

The functioning of the new control method will be verified by checking the output for a few cases. Each case represents a network effect that should be prevented or resolved to maximise the network performance, and there will be checked if the MILP optimisation gives the expected solution. For each case, a fictive network with a PWC demand function and controllers is presented and the functioning of the control method is analysed in three steps:

1. What will happen when the network is not controlled?
2. How should the network be controlled to minimise the cost function?
3. What is the solution of the MILP optimisation?
During the first step the uncontrolled situation is calculated by the PAN model, and will predict what will happen without control. Next there is argued logically which control action(s) can minimise the travel times and there is checked if this matched the solution given by the control method. This will give an impression if the output of the control method is logically valid. Also, comparing the situation without control with the solution of the control method will indicate the value of controlling the network.

At first a simple, trivial, network will be presented to analyse the effects of a capacity drop. Next, the way(s) to prevent blocking back effects in a more extensive network are discussed. In this network coordinated and integrated control is applied to verify whether the method can handle different control measures.

### 6.1.1 The Capacity Drop

Capacity drop occurs when a bottleneck is activated and reduces the maximum outflow of the bottleneck. Appendix 0 shows that preventing a capacity drop will not always guarantee the best network performance, but when a network contains only one bottleneck it seems logical that deactivating this bottleneck will improve the outflow and reduce the queues.

Figure 6.1: Lay out trivial network

Figure 6.1 presents a simple, trivial, network that consists of only three (on-going) links with a bottleneck located between link 2 and link 3. A mainstream controller is implemented at the end of link 1. This network will be used to analyse the effects of a capacity drop on the outflow and the queue formation, since no other effects can occur.

At the origin, traffic flows into the links according to the demand pattern given in Table 6.1.

<table>
<thead>
<tr>
<th>Time (h)</th>
<th>0-1/3</th>
<th>1/3-2/3</th>
<th>2/3-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Demand 1 (veh/h)</td>
<td>6000</td>
<td>5000</td>
<td>3000</td>
</tr>
</tbody>
</table>

Table 6.1: Piecewise constant demand pattern

All the relevant network parameters for this example are displayed below, and will be used to fill the equations of the PAN-model. Note that both bottlenecks can be inactive: \( q_{\text{cap,x}} \) is valid, or active: \( q_{\text{dch,x}} \) is valid.

\[
\begin{align*}
q_{\text{cap}} & \quad \text{Bottleneck capacity} & 5000 & \text{veh/h} \\
q_{\text{dch}} & \quad \text{Bottleneck discharge rate (capacity drop)} & 4000 & \text{veh/h} \\
\rho_{\text{jam,1}} & \quad \text{Jam density link 1} & 200 & \text{veh/km} \\
\rho_{\text{jam,2}} & \quad \text{Jam density link 2} & 200 & \text{veh/km} \\
\rho_{\text{jam,3}} & \quad \text{Jam density link 3} & 200 & \text{veh/km} \\
L_1 & \quad \text{Length link 1} & 2 & \text{km} \\
L_2 & \quad \text{Length link 2} & 2 & \text{km} \\
L_3 & \quad \text{Length link 3} & 2 & \text{km} \\
v_f & \quad \text{Free flow speed} & 120 & \text{km/h} \\
t_k & \quad \text{Time of period} k & 1/3 & \text{h} \\
q_c & \quad \text{Minimum outflow of the active controller} & \geq 1000 & \text{veh/h}
\end{align*}
\]
The effects of controlling a network where a capacity drop can occur will be analysed as well for a situation in which the bottleneck is non-active at start as for a situation where the bottleneck is active at start.

**Preventing the capacity drop**

Without control, the inflow into the bottleneck exceeds the capacity during the first period by 1000 vehicles per hour. The bottleneck will be activated and the outflow will only be \( \frac{4000}{3} \) vehicles during this period. This results in a queue of 667 vehicles at the end of period 1. Due to the presence of this queue, the bottleneck will also be active during the second period, and the queue length will increase to 1000 vehicles. The active bottleneck will also give an outflow of 4000 vehicles per hour during the last period, where the total demand (inflow plus queue) equals 6000 vehicles per hour. This means that a queue of 667 vehicles will remain after one hour, as shown by the grey lines in figure 6.2). The total time spent in queue in this situation is 556 hour.

Preventing the capacity drop from happening will decrease the queue formation, since more vehicles can enter link 3. When the inflow into the bottleneck is kept lower or equal to 5000 vehicles per hour, the bottleneck will stay inactive and the outflow of the network is maximised. This means, that during the first and second period, the controller at the end of link 1 must limit the outflow to 5000 vehicles per hour. Since the demand during the first period is larger than this, a queue of 333 vehicles will built up at link 1 and will remain during the second period. Due to the lower demand during the third period, the queue can be resolved within the hour. The whole period, there will not be a queue at link 2, since all the vehicles that enter link 2 can leave the link in free flow.

Now there is sketched what would happen if the network is not controlled and there is analysed what should be done to improve the situation, there is checked whether the control method gives the same solution. The network is implemented in MATLAB as a MILP problem and optimised by the use of TOMLAB cplex. This gives the following values for the controller outflow:

- \( q_c(1) = 5000 \) vehicles per hour
- \( q_c(2) = 5000 \) vehicles per hour
- \( q_c(3) = 4000 \) vehicles per hour

![Figure 6.2: Queue development](image)

The queue lengths given by the control method are displayed by the black lines in figure 6.2, which match the expectations as identified before. The surface covered by the black lines is much smaller than the surface under the grey lines, which can be verified by comparing the values for the cost function: in the controlled situation the solution for the cost function is decreased to 167 hour, which is only 30% of the original time spent in queue.
In this example an active bottleneck is assumed at start, which is modelled by adding a queue of 100 vehicles on link 2 at the start of the first period. In the uncontrolled situation the queues on link 2 and link 1 develop just as in the previous example, although the queue on link 1 is 100 vehicles larger. This is shown by the grey line in figure 6.3. The total time spent in queue in this situation is 750 hour.

Figure 6.3: Queue development with active bottleneck at start

Also in this case the network performance could be improved by keeping the inflow into the bottleneck under the capacity, but at first the bottleneck must be deactivated by resolving the start queue on link 2. In practice the bottleneck should be deactivated at fast as possible, but due to the use of periods in the PAN model the queue must be resolved within one period to increase the outflow during the second period. The queue can be resolved by limiting the outflow of link 1 to \( \frac{4000}{3} - 100 = \frac{3700}{3} \) vehicles during the first period. This means that a queue of \( \frac{2300}{3} \) vehicles will be built up during the first period, and will remain this size during the second period. After one hour \( \frac{4000}{3} + \frac{5000}{3} + \frac{5000}{3} = \frac{14000}{3} \) vehicles can be flown out of the network due to the active bottleneck at start. Since the total demand for this hour is \( \frac{14000}{3} \) vehicles plus the 100 vehicles in queue, 100 vehicles remain in queue after one hour.

The output of the MILP optimisation determines the same queue development as shown by the black lines in figure 6.3, by using the following controller settings:

- \( q_c(1) = 3700 \) vehicles per hour
- \( q_c(2) = 5000 \) vehicles per hour
- \( q_c(3) = 5000 \) vehicles per hour

The total time spent in queue in this situation is 456 hour, which is 68% of the solution for the uncontrolled situation.

6.1.2 Blocking back effects

Determining the blocking back effects requires a more extensive network than the network of the previous cases. An off-ramp and an on-ramp are added to the network to identify the effects of spillback on the turning vehicles and the underlying road network. Since the control method must be able to deal with coordinated and integrated control, ramp metering is implemented at the on-ramp. This results in a network that covers the network effects due to capacity drop and spillback, and is suitable for coordinated and integrated control (figure 6.4).
Figure 6.4: Lay-out non trivial network

Table 6.2 gives the PWC demand pattern for both origins:

<table>
<thead>
<tr>
<th>Time (h)</th>
<th>0-1/3</th>
<th>1/3-2/3</th>
<th>2/3-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Demand 1 (veh/h)</td>
<td>5000</td>
<td>5000</td>
<td>4000</td>
</tr>
<tr>
<td>Demand 2 (veh/h)</td>
<td>2000</td>
<td>2000</td>
<td>1000</td>
</tr>
</tbody>
</table>

Table 6.2: Traffic demand validation network

All the relevant network parameters for this example are given below, and will be used to fill the equations of the PAN-mode. Note that both bottlenecks can be inactive: than \( q_{\text{cap},x} \) is valid, or active: than \( q_{\text{dch},x} \) is valid.

\[
\begin{align*}
q_{\text{cap}} & \quad \text{Bottleneck capacity} & 6000 & \text{veh/h} \\
q_{\text{dch}} & \quad \text{Bottleneck discharge rate (capacity drop)} & 4800 & \text{veh/h} \\
\rho_{\text{jam},1} & \quad \text{Jam density link 1} & 200 & \text{veh/km} \\
\rho_{\text{jam},2} & \quad \text{Jam density link 2} & 200 & \text{veh/km} \\
L_1 & \quad \text{Length link 1} & 2 & \text{km} \\
L_2 & \quad \text{Length link 2} & 2 & \text{km} \\
t_k & \quad \text{Time of period }k & 1/3 & \text{h} \\
q_{c1} & \quad \text{Minimum outflow of active controller 1} & 3000 & \text{veh/h} \\
q_{c2} & \quad \text{Minimum outflow of active controller 2} & 1000 & \text{veh/h} \\
\beta & \quad \text{Turn fraction (to destination 1)} & 0.1 & \\
\end{align*}
\]

This section analyses at first the way in which the network should be controlled to prevent blocking back effects. Next there is looked what will change when a maximum queue length on the on-ramp is used to prevent spillback of congestion to the underlying network.

Preventing blockage of off-ramp

At the network of figure 6.4, two network effects can occur: capacity drop at the bottleneck, and blockage of the off-ramp by spillback of the congestion on link 3. Without control, the inflow into the bottleneck exceeds the capacity during the first period and the capacity will drop with 20%. A queue is built up, and causes a remaining capacity drop during the second and third period. Figure 6.5 shows that the outflow of link 5 is reduced to the discharge capacity of 4800 vehicles per hour, but also the outflow of link 2 is lower than the demand. Spillback of the queue on link 3 hinders the outflow of link 1 into link 2. In this uncontrolled situation queue arise at link 3 and link 4 and will not be resolved within the hour. This give a total time spent in queue of 816 hour.

The cost function can be lowered by controlling the network in such a way that both the capacity drop and the blocking back effects are prevented. This can be done by ensuring an inactive bottleneck and by preventing a queue building up at link 3. In this case, preventing a queue on link 3 is also required to keep the bottleneck inactive.
By keeping the inflow into the bottleneck lower or equal to the bottleneck capacity the capacity drop is prevented, but this can be done in two ways:

1. Limiting the outflow on link 1 by controller 1
2. Limiting the outflow of link 4 by controller 2

The first action will limit the outflow of the network through link 2. Therefore, action 2 will give better results in terms of outflow of the network and the number of vehicles in queue.

Figure 6.5: Blocking back effects

The output of the control method indeed indicates action 1 as the best solution. The controller at the end of link 1 will be inactive, and the flow is only limited by controller 2:

- $q_{c2}(1) = 1500$ vehicles per hour
- $q_{c2}(2) = 1500$ vehicles per hour
- $q_{c2}(3) = 2000$ vehicles per hour

Figure 6.5 shows the outflow of the network as given by the control method, compared with the uncontrolled situation. The outflow of both link 2 and link 5 will be higher in the controlled situation, since the bottleneck is inactive and the inflow into link 2 is not hindered by spillback. In the uncontrolled situation, link 3 gets fully congested within the first period and already than spills back to link 1 (figure 6.6). The longer the queue on link 1 becomes the more the outflow of link 2 is hindered, which can be seen in the relation between both figures in this section.

Figure 6.6: Queue development when preventing blocking back effects

Controlling the network in this case gives reduces the total time spent in queue with almost 83% to only 139 hour. The most beneficial action in this situation is preventing the capacity drop, but the higher the turning rate will be in a network, the more can be gained with preventing blocking back effects.
The last case adds a constraint to the model concerning the maximal queue length on link 4, to prevent disturbance of the connecting (urban) network. There is chosen for a maximum queue length of 100 vehicles on link 4, which is exceeded in the previous section.

Because of this constraint, controller 2 should provide an outflow of at least $D_2 \epsilon_k - 100$ vehicles. Without using controller 1 the inflow into the bottleneck will be higher than the capacity and the bottleneck will become active. Therefore, the two controllers must work together to prevent the capacity drop as well as spillback to the underlying network. It is expected that the best performance will be given when controller 2 limits the inflow into the bottleneck as much as possible (given the constraint to the queue length), so the outflow of the network through link 2 will be limited as less as possible.

![Figure 6.7: Queue development when preventing spillback to underlying network](image)

Figure 6.7 compares the output values given by the control method for the situation with and without constraints for the maximum queue length on link 4. Where in the previous situation only a queue was built up at link 4, this queue is now limited to 100 vehicles and another queue is built up at link 1 due to the activation of controller 1:

- $q_{c1}(1) = 4778$ vehicles per hour
- $q_{c1}(2) = 4444$ vehicles per hour
- $q_{c1}(3) = 4778$ vehicles per hour
- $q_{c2}(1) = 1700$ vehicles per hour
- $q_{c2}(2) = 2000$ vehicles per hour
- $q_{c2}(3) = 1300$ vehicles per hour

The solution of the cost function is 149 hour when using these controller settings. This a bit higher than the unconstrained situation, due to the fact that controller 1 limits the outflow of the network through link 2.

6.1.3 Overview

The four cases that are analysed in this paragraph show that the control method behaves as expected to be logically valid. The control method focusses on preventing the network effects that influence the cost function most, and is proven to be suitable for coordinated and integrated control in the last case.

To ensure the correctness of the transformation of the PAN model into a MILP problem, the output given by the control method is compared with the values given by the non-linear PAN model (when using the same controller settings): this gave for all four cases the same output. This means that the transformation is performed correctly in these cases.
The improvements, in terms of total travel time spent in the network, due to the control method are pictured in table 6.3. This shows that the application of the control method can achieve large reductions of the total time spent in the network. The smallest improvement is seen in the situation where the bottleneck is active at start; this can be explained due to the fact that the bottleneck is active at start: no improvements in outflow can be achieved during the first period.

<table>
<thead>
<tr>
<th></th>
<th>Uncontrolled situation</th>
<th>Controlled situation</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preventing the capacity drop</td>
<td>556</td>
<td>167</td>
<td>-70%</td>
</tr>
<tr>
<td>Resolving the capacity drop</td>
<td>622</td>
<td>456</td>
<td>-27%</td>
</tr>
<tr>
<td>Preventing blocking back</td>
<td>816</td>
<td>139</td>
<td>-83%</td>
</tr>
<tr>
<td>Preventing spillback</td>
<td>816</td>
<td>149</td>
<td>-82%</td>
</tr>
</tbody>
</table>

Table 6.3: Comparison total time spent in the network (h)

### 6.2 Computation time operations PAN model

The value of the new control method over other methods will be covered in the computation time needed for the optimisation and therefore the network size to which the method can be applied. Based on literature this paragraph will research the relation between the number of variables and the computation complexity of MILP problems, and in particularly the complexity of the new control method. This information will give more insight in the computation complexity and the consequences this will have on the network size of the traffic model.

In the previous paragraph the control method is applied to small sized networks, which need only short computation time to obtain the solution. Table 6.4 gives an overview of the computation time and number of variables and constraints for these cases.

<table>
<thead>
<tr>
<th></th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU time (sec)</td>
<td>0.05369</td>
<td>0.05622</td>
<td>0.11634</td>
<td>0.12054</td>
</tr>
<tr>
<td># Binary variables</td>
<td>10</td>
<td>10</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td># Constraints</td>
<td>70</td>
<td>70</td>
<td>106</td>
<td>106</td>
</tr>
</tbody>
</table>

Table 6.4: CPU time and number of variables/constraints for cases

Since it is desired to know how much the network size can be enlarged and still be suitable for real time control, the relation between the network size and the computation time must be known. Table 6.4 shows a positive relation between the number of binary variables and constraints and the CPU time, but this does not state anything about larger networks.

The study of Till et al. (2004), defines that the practical complexity of the CPU for MILP problems varies from \(O(dm^3)\) to \(O(2^d m^3)\), the worst case scenario gives \(O(2^{m+d})\). Here \(d\) indicates the number of binary variables and \(m\) the number of constraints. This shows that in worst case scenario the complexity of a MILP problem grows exponentially, same as non-linear programming problems. Here raises the question in which situations solving a MILP problem will require less computation time than when non-linear programming is used? The answer can be found in the formulation of the practical complexity: this depends only on the number of binary variables and number of constraints. The notation \(O(2^d m^3)\) also indicates that the number of binary variables has a larger influence on the computation complexity than the number of constraints.

For the PAN model, it is important to indicate which factors in the model influence \(d\) and \(m\). The previous chapter has shown that the largest share in the constraints in the model is formulated when a non-linear equation is transformed into a linear, constrained equation. In this situation also the binary variables occur. This means that the number of non-linear equations in the PAN model determines the number of \(d\) and \(m\).
The network lay-out determines the number of non-linear equations in the PAN model. Chapter 4 has shown that each link that is situated between two other links gives a non-linear equation due to the inclusion of spillback. When a bottleneck is presented at the end of a link, this non-linear equation is extended and returns more constraints and binary variables during the transformation. Also the choice for the duration of a time period influence the number of \( d \) and \( m \): changing the time period from 20 minutes to (for example) 1 minute will multiply the number of variables and the number of constraints with 20.

Chapter 5 and appendix C show that two links, connected by a bottleneck, produce 10 binary variables \( z_i \) and 10 auxiliary variables \( z_i \) per time period. Each binary variable is covered in 2 or 3 constraints, each auxiliary variable in 4 constraints. For a network of \( x \) on-going links and \( y \) bottlenecks, this results per period in:

- \( 10y \) to \( 10x \) binary variables
- \((2 \text{ or } 3)(10y \text{ to } 10x)\) constraints for the binary variables
- \(4(10y \text{ to } 10x)\) constraints for the auxiliary variables

In worst case, each extra link in a network will result in 70 extra constraints and 10 extra binary variables per time period. The effect of enlarging the network size on the computation time cannot be given or predicted by this information, but it does indicate which factors do influence the computation complexity. This means that traffic networks which can be represented by the PAN model in few non-linear equations would probably return a smaller CPU time than networks which contain many non-linear equations, also when the first network is larger in size.

### 6.3 Overview

The first paragraph has shown that the output of the simplified, reformulated control problem seems to be logically valid. Also, the values of the network parameters given by the MILP optimisation match the output of the original PAN model when using the controller settings as input; this ensures the correct transformation of the PAN model towards the constrained linear model in these cases.

The advantage of solving a MILP problem over a nonlinear problem can be identified in terms of computation time needed. Literature has shown that the growth function of the computation complexity of MILP problems is, just as for NLP problems, exponential but also that the CPU time needed depends mostly on the number of binary variables and constraints. The number of binary variables and constraints will be determined by the number of non-linear equations in the PAN model. Based on this information it is expected that solving a MILP problem instead of a NLP problem will give advantage in terms of CPU time when the networks can be represented by few non-linear equations. This expectation can be verified when further research will implement the control method to a larger diversity of control problems.

At last, a note must be made about the effects of using the MILP approach as MPC controller. Using feedback means that at every time period there must be checked whether the predicted solution still fits, which result in extra calculations. This will further limit the maximum size of the network but will also give extra advantage over the non-linear approach, since there can be gained in computation time every loop.
7. CONCLUSION AND RECOMMENDATIONS

This research has started with an objective: to develop a control method for large traffic networks that can optimise the network performance in real time while there is aimed at realistic representation of the traffic situation.

Whether this objective is achieved will be stated in this chapter, by answering all the research questions and discussing all the findings of the research. This is followed by a critical evaluation of the method: the strengths and weaknesses are mentioned in the third paragraph. This leads to a couple of recommendations for further research.

7.1 ANSWERING THE RESEARCH QUESTIONS

A new control method is developed in this research that should be able to optimise large freeway networks real time. Answering the first research question shows the control features that must be included in the control method to reach the optimal solution network-wide:

What are the requirements for a real time and network-wide control method?
The best performance network-wide, instead of local, can be reached when the DTM measures are coordinated and integrated. The collaboration of different type of DTM measures can improve the traffic situation over the use of a single type of measure, since more control options are available. When the control method will be applied to a traffic network it is important to represent this network as realistic as possible. This means that the control method must be able to handle the dynamics of the traffic networks: the time varying demand and the traffic assignment. Also, the network effects of freeways (capacity drop, blocking back and shockwaves) must be included in the control method. Besides these features, the use of feedback control can further improve the network performance: the solution given by the control method will be checked every time step and adjusted when needed. In this way, any unexpected behaviour can be identified and the earlier calculated settings for the DTM measures can be corrected.

Based on these requirements, a control approach is chosen. An approach that uses feedback control and can be used for a network-wide control method is Model Predictive Control (MPC), which will give the best approximation of the optimal solution but also has a high computation complexity. To ensure the possibility of real time control this computation complexity must be reduced, which gives the second research question.

Further, the traffic situation must be represented by a traffic model which can represent all the required network effects (blocking back effects and capacity drop) and can deal with the DTM measures ramp metering and mainstream control. Also, this model must match the control approach.

How can the computation time be reduced?
The computation complexity, and herewith the computation time, of MPC can be reduced in two steps. At first, the situation (represented by a traffic model) must be simplified. This means that the traffic model which is used to represent the traffic network and which gives input to the control method must be as simple as possible. The next step is transforming the non-linear control problem into a Mixed Integer Linear Programming (MILP) problem by the use of logic rules and additional constraints. This can speed up the optimisation at least 100-400 times.

The new control method is developed based on the answers of the previous two research questions. To determine the value of the control method, the following research question is answered:
What is the advantage of the new control method over other control methods in terms of computation time and approximation of the optimal solution?

The computation complexity of the new control method has a growth function which is exponential in worst case, depending on the number of constraints and binary variables in worst case. Non-linear programming problems also have a complexity which grows exponentially, this means that the advantage of the new control method over non-linear control approaches depends on the network where the method will be applied. When a network can be covered by few non-linear equations, less binary variables and constraints are formulated in the control problem than for a network with many non-linear equations. In the first situation it is expected that the new control method can return smaller CPU times than non-linear control methods. For other cases no statements can be made based on this research: this should be further researched.

Whether the control method returns the optimal solution for the real network could not be stated in this research. However, verification of the control method shows that it gives an output which is logically valid in a few cases.

7.2 Research Conclusions

The findings of the research that are (partly) mentioned in the previous paragraph lead to the conclusion of this research: how can they be explained and what do they mean for the research objective?

There is shown that two main goals of the control method: the possibility to control large network real time, and maximising the network performance (by approximating the optimal solution as much as possible), are contradictory. The real time control requires short computation times, and therefore a low computation complexity, while the features to improve the solution result in a higher computation complexity.

This research has developed an approach to maximise the combination of both by the use of efficient MPC. A few concessions are done regarding the way in which the traffic network is represented in the PAN model, to make this model piecewise affine and suitable for transformation into a MLD model. The MLD model can be solved as a MILP problem, which can be solved in less CPU time than NLP problems when the number of binary variables and constraints is small. This means that value of the new control method over non-linear control methods can be expected when the control problem contains less binary variables and constraints, for other cases it cannot be determined.

At the end of the research, there can be concluded that a new control method is developed which is suitable for the coordinated and integrated real time control of larger freeway networks than other feedback control methods. Although it is expected that the control method will not always return the optimal solution (due to the simplifications), the gain in computation time adds a lot of value to the method. Besides, the most other control methods are also not able to return the optimal solution. There can be stated that the research objective is reached but many options for improvements are left, which will be discussed in the next two paragraphs.
7.3 Discussion

This research is a starting point for the network-wide and real time control of large traffic networks. The new control method offers a reduction in computation complexity which makes it suitable for larger networks than other feedback control method. However, the control method also has a few drawbacks which were mentioned through the report and will be recalled below. This will be the base for the recommendations in the next paragraph.

- The PAN model cannot represent networks with multiple routes due to the use of time periods instead of a time horizon. This means that the control method cannot be applied to networks with multiple routes.
- The network state in the PAN model cannot change during period $k$: bottlenecks cannot switch from active to inactive (or vice versa). This can result in higher or lower outflows of the bottleneck, which will give wrong information to the control method and possibly non-optimal controller settings for the next period. However, this effect can be reduced by the use of very small time periods.
- The PAN model indicates that congestion downstream does affect the link upstream until the jam density is reached. This means that the PAN model can determine a higher outflow of a link. And this can also result in non-optimal controller settings when the method is applied to reality or another traffic model.
- The advantages of the control method over other control methods could not be quantified in terms of computation time, network size and solution since an extended validation is not performed.
- The control method in this form is not suitable yet for application to another traffic model or to a real network, since the MPC coupling is not yet designed.

7.4 Recommendations for further research

Evaluation of this research has shown that there are many possibilities left to expand and improve the control method. This paragraph elaborates on the three most important subjects for further research, with as goal to make the control method suitable for application to real traffic network.

Expansion PAN model

The PAN model can be further extended to represent more network effects, to make the control method suitable for more network types and for more DTM measures. The following recommendations are formulated for expansion of the PAN model:

- Adding the effects of shockwaves to the PAN model will solve the drawback of the model that the situation downstream will not influence a link until the link downstream reaches the jam density. It will also represent the capacity drop in jams.
- Translating the use of time periods in a time horizon which indicates the time at which effects occur. This will make the PAN model suitable for networks with multiple routes.
- Implementing more DTM measures, such as route guidance and variable speed limits, will increase the possibilities to control the traffic network. This can lead to improvement of the network performance. However, the effects of these measures must first be captured mathematically in the PAN model.

When expanding the PAN model with these features it is of importance for the control method that the model is kept piecewise affine, otherwise the transformation to a MLD model is not possible anymore.
Extended verification and validation
The PAN model as well as the control method is only briefly validated in this research, which did not give all the information about the robustness and the advantages over other control methods. Therefore it is recommended to perform a couple of extra validation steps:

- Verification of the PAN model by comparing the model results with the output of other model and road data. This will give information about the extent to which the PAN model represents the real traffic situation, and the model can be calibrated based on the data.
- The relation between the network size and the computation time needed must be further researched to give information about the scope of the control method. This can be done by applying the control method to increasing network sizes, and to compare the CPU time given. This will indicate the maximum network size (in terms of constraints and binary variables) that can be controlled real time. When the same step is performed for non-linear approaches, statements can be made about the difference in network size that can be controlled.
- Researching the possibilities of further simplification of PAN model and the effects this will have on the solution and the CPU time. With the simplification, the number of constraints and binary variables can be reduced, which should decrease the CPU time according to the literature. It is interesting to know what would be the effects of the simplifications on the solution given by the control method for discussion about the trade-off between solution and computation time.

Implementation of control method as MPC controller
Before the control method can be applied to real traffic networks, a feedback coupling between the control method and the traffic network must be designed. The following steps concerning this coupling are recommended:

- The transformation from PAN model to MLD model and the construction of the MILP input by cplex can be automated. In this research it is done by hand, but it is possible to program this transformation in such a way that the MILP input will be given based on the network lay-out. This will avoid human errors and will reduce the time that is needed to construct the MILP problems, especially when the control method is applied to large networks.
- Couple the control method first as MPC controller to other traffic models. This is useful for two goals: at first, there can be checked if the extension from control method to MPC controller is working properly. Further, this coupling can be used to validate the control method by comparing the functioning as MPC controller with traditional MPC control.
- For the coupling between the control method and real networks, the sensors that measure the network performance at freeways must be used as input for the control method. These sensors must measure the flows and queue lengths, since these parameters are used by the PAN model to determine the traffic state. It is of importance that the correct information is directly transferred: every delay in information will give a delay of the controller settings, which can worsen the network performance. Therefore it is recommended that fast and reliable data transmission between the sensors and the control method is ensured.
REFERENCES


Verhoef, E. e. (2004). *Vormgeving en effecten van prijsbeleid op de weg*. Amsterdam: SEO.


A. APPENDIX: NETWORK EFFECTS

To support the methodology which states the importance of a network wide approach for the control of DTM measures, a few examples are given to show the network-wide effects of DTM measures. At first, the effects of single DTM measures on the outflow are determined. The second section will give examples of situations in which the coordinated use of measures can improve the traffic situation even more. This information will be used in the third section to explore the effects of DTM measures in bigger networks.

A.1 SINGLE DTM MEASURES

The effects of individual use of ramp metering, DRG and variable speed limits are determined below by three examples. These examples elaborate on situations where DTM measures can be used to deactivate bottlenecks and improve the outflow hereby.

A.1.1 RAMP METERING

Consider a freeway trajectory where an off-ramp is followed by an on-ramp (figure a.1). A bottleneck is situated at the location where the on-ramp and the main road are connected. When the traffic demand at the main road plus the traffic demand at the on-ramp exceed the capacity of the bottleneck, traffic jams are formed and a capacity drop of the bottleneck occurs. This results in spillback on both the main road and the on-ramp and blocking back effects can be identified when the jam on the main road reaches the off-ramp. Ramp metering can be used to reduce the traffic demand at the bottleneck, by allowing fewer cars from the on-ramp enter the main road.

In this example a static situation is assumed, with the following parameters:

- Inflow main road (upstream the off ramp)= 2000 vehicles/hour
- Inflow on-ramp= 500 vehicles/hour
- Turning rate off-ramp= 20%
Figure A.2 shows the fundamental diagram of this situation, where the maximum capacity is 2000 vehicles/hour, the capacity drop 10%, the critical density 20 vehicles/km and the jam density 100 vehicles/km. This gives a free flow speed of 100 km/hour.

Without ramp metering, the traffic demand at the bottleneck is exceeding the capacity by 100 vehicles/hour and will activate the bottleneck. This causes the capacity drop, what means that the traffic demand becomes 300 vehicles/hour above the capacity of the bottleneck. At the start, the density at the main road between the off-ramp and on-ramp is 1600/100= 16 vehicles/km but this will increase with 5 vehicles/km every minute what will result in stand still situation where the off-ramp is blocked. In this situation, the traffic demand is 2500 vehicles/hour, but the outflow is only 1800 vehicles/hour.

When ramp metering is active, the inflow from the bottleneck can be reduced to 400 vehicles/hour, this prevents exceeding the capacity of the bottleneck. Congestion will not occur, and the outflow of the network will be 2400 vehicles/hour. This is much higher than the situation without ramp metering.

A.1.2 Dynamic Route Guidance

Consider an origin X and a destination Y which are connected by 2 routes, both containing a bottleneck (figure a.3). Different situation can occur, depending on the traffic demand. In this example both bottlenecks are active due to the past situation and the following parameters are relevant:

- Inflow X= 1900 vehicles/hour
- Maximum capacity main road= 2000 vehicles/hour
- Maximum capacity at B1 and B2= 1000 vehicles/hour
- Capacity drop at B1 and B2= 10%
- Turning rate route 1= 50%

![Figure A.3: Example DRIP](image)

Without measures, both routes will be congested and the outflow at destination Y will be only 1800 vehicles/hour. When a DRIP is activated, the bottlenecks can be deactivated one by one. At first the DRG reduces the inflow at route 1 by sending more vehicles over route 2. When bottleneck 1 is deactivated, there can be send more vehicles over route 2 what will reduce the flow on route 2 such that this bottleneck will be deactivated as well.
A.1.3 **Variable Speed Limits**

This example consists only of a main road and an onramp (figure a.4). When the traffic demand from both directions is higher than the capacity of the bottleneck, congestion occurs. Instead of using ramp metering, as shown in example 1, it is also possible to reduce the inflow from the main road by a lower speed limit.

![Figure A.4: Example DSL](image)

The effects of a lower speed limit (80 km/hour) on the fundamental diagram are shown in figure a.5: the critical density and the jam density remain the same value, but the slope of the free flow speed decreases. This results in a maximum flow of 1600 vehicles/hour at the critical density.

![Figure A.5: Fundamental diagram of example 3, with a lower speed limit](image)

In theory the new speed limit will result in a lower flow at the bottleneck, what will prevent the activation of the bottleneck. In practice these effects can be different, and there must be mentioned that the new value for the flow at capacity can create a new bottleneck when the traffic demand is higher than the new capacity. Papageorgiou mentions that VSL are only effective just before a bottleneck, other evaluations of pilot projects have shown that in most cases lower speed limits result in a lower outflow. In theory, a lower speed limit before the bottleneck and a higher speed limit after the bottleneck can improve the outflow, but the estimation of effects of frequently adjusting the speed limit on the road user is difficult.
A.2 Coordinated use of DTM measures

The previous section has shown the effect of the use of single DTM measures at bottlenecks. In some situations, single measures will not be sufficient to deactivate a bottleneck or the traffic situation in a network can be improved further by the implementation of combined measures. This paragraph will give two examples to illustrate this.

A.2.1 VSL and ramp metering

Example 1 gives a situation with an on-ramp, where ramp metering is implemented. This example has shown that reducing the inflow from the on-ramp by 100 vehicles/hour will deactivate the bottleneck. It will also cause a queue on the on-ramp, which might spill back to another network and cause problems there. To prevent this, a maximum value for the queue length can be stated which will influence the minimum flow passing the ramp metering. In most cases, the restrictions for the queue length will result in a higher inflow from the on-ramp than desired to deactivate the bottleneck. In these situations, an additional measure can be used (see figure A.6).

Figure A.6: Example coordinated use DSL and ramp metering

The implementation of ramp metering as well as dynamic speed limits can reduce the flow from both the main road and the on-ramp. This will cause lower effects upstream the measures than the use of single measures, since the controller settings for both measures will be less fierce in this situation. Hegyi has shown that the use of both measures result in a larger set of controller settings to deactivate the bottleneck. For a network-wide approach this is desired: there are more options, what can result in a better network performance during the optimisation phase.

A.2.2 DRG and ramp metering

Example 2 gives a small network of 2 routes, with active bottlenecks and DRG to control the route choice. For the chosen parameters in this example, DRG can deactivate both bottlenecks. But when considering the same parameters but a higher traffic demand, for example 2000 vehicles/hour, the DRIP will not be able to deactivate both bottlenecks.

Figure A.7: Example coordinated use DRIP and ramp metering

The bottleneck on route 1 can be deactivated by sending temporary more traffic over route 2 (900 vehicles/hour versus 1100 vehicles/hour). After deactivation route 1 can process 1000 vehicles/hour, what results in also 1000 vehicles on route 2: this is too high to deactivate the bottleneck. When the bottleneck of route 2 consists of an on-ramp, ramp metering can be implemented to reduce the inflow temporary what can be sufficient to deactivate the bottleneck. At this point, the
ramp metering is no longer needed since the capacity of route 2 will also be upgraded to 1000 vehicles/hour.

A.3 NETWORK-WIDE APPROACH

Previous examples have elaborated on the deactivation of bottlenecks by the use of (coordinated) DTM measures. One of the effects of this deactivation is the higher outflow of a bottleneck, what can affect the bottlenecks downstream.

![Figure A.8: Coupled bottlenecks on corridor](image)

In figure A.8, a freeway with 3 bottlenecks is shown to illustrate the coherence between bottlenecks. When bottleneck 1 is in initial situation active, bottlenecks 2 and 3 are inactive and due to DTM measures bottleneck 1 is deactivated. This will result in a higher outflow of this bottleneck, which can cause activation of the second bottleneck. When this situation continues, spillback from bottleneck 2 can also cause reactivation of the first bottleneck. To improve the outflow of this corridor, this example demonstrates the importance of adjusting the controller settings of a bottleneck to the traffic condition downstream.

Another example is given in figure A.9, where two corridors are connected by a link which is the off-ramp of corridor 1 and the on-ramp of corridor 2. This situation illustrates the complexity of traffic networks, since all the bottlenecks can influence the others.

![Figure A.9: Coupled bottlenecks on two corridors](image)

In this example, the best network performance in terms of total outflow might be achieved by the activation of bottleneck 1 at corridor 2. This will give a lower inflow at the bottlenecks 2 and 5 which can prevent these from activation what results in a continuous outflow.

When bottleneck 4 is following an important road section, creating congestion here can be undesired and it could be a better option to activate bottleneck 1 instead.

These two examples show the complexity of networks and the coherence among the bottlenecks. For small networks it will be possible to determine the required DTM measures by hand, for larger ones programming is needed to reach optimisation of the network performance.
B. Appendix: Link Transmission Model

In chapter 3, traffic models are ranked based on their features and indicates the Link Transmission Model (LTM) as the best choice overall for this research. Further exploration has shown that the model was not suitable for this research, which will be shown in this appendix.

At first the model and its equations will be described, followed by a proposal for extension of the model to meet all the model requirements as identified in the methodology. At last the reasons for rejection of the LTM for this research are given.

B.1 Model description

The link transmission model is developed as part of the promotion research of Isaak Yperman and has the following characteristics (Yperman I., 2007):

- The network consists of homogeneous links and nodes between the links
- Traffic flows move over the link according to the first order kinematic wave theory
- The forming and dissolving of queues fit the reality better than state-of-art traffic models
- Traffic flows at bottlenecks behave according to the queuing theory
- Vehicles can be disaggregated over the available routes
- The LTM algorithm is efficient: traffic flows in large networks can be simulated in little computation time

The mathematics behind the flow propagation will first be explained briefly to give an impression of the model. Since it is possible to use the LTM for dynamic network loading, the second part will deepen into the options for route set generation and the way in which this can be used.

B.1.1 Flow propagation

In the LTM, traffic propagates on a link as assumed in kinematic wave theory, which is based on the conservation of vehicles concept:

\[ q(x, t) = \frac{\partial N(x, t)}{\partial t} \]  \hspace{1cm} B.1

\[ \rho(x, t) = \frac{\partial N(x, t)}{\partial x} \] \hspace{1cm} B.2

\[ \frac{\partial q(x, t)}{\partial x} + \frac{\partial \rho(x, t)}{\partial t} = 0 \] \hspace{1cm} B.3

Here \( q \) represent the flow in vehicles per hour, \( \rho \) the density in vehicles per kilometre, \( \partial N \) the change in vehicles (over time \( t \) or over distance \( x \)).

Equation B.3 is the well-known form of the conservation law, and gives the Calculus theory:

\[ N(x_2, t_2) - N(x_1, t_1) = \int q dt - \rho dx \] \hspace{1cm} B.4

The LTM uses the kinematic wave theory to describe the situation in the network at time \( t \) in terms of cumulative vehicle numbers that have passed the beginning of a link \( i \): \( N(x_i^0, t) \) and the end of a link \( i \): \( N(x_i^0, t) \).

The transmission of vehicles between links is done by calculating the number of vehicles that want to leave the link: the sending flow \( S_i(t) \) and the number of vehicles that can enter the link downstream: the receiving flow \( R_i(t) \).
The sending flow is defined as the maximum amount of vehicles that can leave a link to the link downstream during a time period $\Delta t$:

$$ S_i(t) = \min \left( \left( N \left( x_i^0, t + \Delta t - \frac{L_i}{w_f}, t \right) - N(x_i^0, t) \right), q_{M,i} \Delta t \right) $$  

Equation B.5 means that the number of vehicles that can leave a link depends on the number of vehicles that have reached the end of the link (based on the time that is needed to move over the link, where $L_i$ is the link length and $w_f$ the free flow speed) and the capacity of the link ($q_{M,i} \Delta t$). For the receiving flow, a similar equation can be given. The receiving flow is also bounded by the capacity of the link ($q_{M,i} \Delta t$), and the fact that a vehicle can only enter the link when there is space left. This is calculated by the presence of shockwaves ($w_f$), the maximum density of the link ($\rho_{jam(i)}$) and the vehicles that have entered the link before. This gives equation B.6:

$$ R_j = \min \left( \left( N \left( x_j^0, t + \Delta t - \frac{L_j}{w_j} \right) + \rho_{jam}, \Delta t - N(x_j^0, t) \right), q_{M,j} \Delta t \right) $$  

The effects of merging and diverging is also adopted by the model, but will not be discussed here.

### B.1.2 Link Transmission Algorithm

The LTM model as described in the research of (Yperman I., 2007) is intended for dynamic network loading, meaning that it contains an algorithm to update the route choice of vehicles according to the traffic situation in the network. The steps of this algorithm are given below:

1. Determine the sending $S_i(t)$ and receiving $R_j(t)$ flows for each link
2. Determine the transition flow $G_{ij}(t) = \min( S_i(t), R_j(t))$ between links
3. Update the vehicle numbers ($N(x_i^0, t)$ and $N(x_i^1, t)$) per link per route

This algorithm must be removed from the model since influencing the route choice is not the only instrument to optimise the network performance and only one cost function is allowed to secure a simple optimisation problem. The possibility to remove this algorithm without destroying the functioning of the model is researched in the next section.

### B.2 Extension

The Link Transmission Model in its original form is not suitable for the research because it contains an algorithm and it does not include the capacity drop or DTM measures. In this paragraph the model is extended to make it suitable for this research.

#### B.1.1 Removing Algorithm

Changing the routes of the vehicles is included in step 3 of the solution algorithm, and is the part that must be removed from the model. When doing so, correct transition of vehicles between links must be secured. Also, there must be found a way to include the routes in the model. When assuming fixed routes, an easy option is the use of turning rates. Removing the algorithm can be achieved by changing the equations of the LTM in such a way that the sending flow of a link is also bounded by the receiving flow of the link downstream the time step before:

$$ S_i(t + \Delta t) = \min \left( \left( N \left( x_i^0, t + \Delta t - \frac{L_i}{w_f}, t \right) - N(x_i^1, t) \right), q_{M,i} \Delta t, R_j(t) \right) $$  

---

Page 68  
Network-wide control of freeway networks
Besides adding an extra bound to the sending flow, the queue formation due to the difference between the sending flow and the receiving flow must also be taken into account. This is represented by the difference between the numbers of vehicles that reach the end of a link \(N \left( x_i^0, t + \Delta t - \frac{L_i}{v_{f,i}} \right)\) at time \(t + \Delta t\) and the number of vehicles that can actually leave the link \(R_j(t)\).

\[
N(x_i^t, t + \Delta t) = N(x_i^0, t + \Delta t) + N \left( x_i^0, t + \Delta t - \frac{L_i}{v_{f,i}} \right)
\]

\[
w(x_i, t + \Delta t) = w(x_i, t) + N \left( x_i^0, t + \Delta t - \frac{L_i}{v_{f,i}} \right) - R_j(t)
\]

**B.1.2 Adding the Capacity Drop**

It is considered that the capacity drop can best be included in the constraints on the receiving flows (equation 6) because an active bottleneck decreases the outflow of the bottleneck into the link downstream. After some discussions it is assumed that changing the capacity of the link downstream the bottleneck according to equation B.10 will give a good approximation of the capacity drop.

\[
q_{M,j}^i = \begin{cases} 
q_{in} & \text{when } q_{in} \leq q_{M,j} \\
q_{dch} & \text{when } q_{in} > q_{M,j}
\end{cases}
\]

There must be noted that adding these conditions affects the capacity of the entire link, but since the link is homogeneous the inflow will always equal the flow on the link.

Another comment is the fact that the capacity drop is originally triggered when the critical density is exceeded and not by an inflow which is higher than the capacity. This might lead to incorrect reactions what will be evaluated when this extension is implemented in the model.

At last, using this representation of the capacity drop can give infinite fast shockwaves as pictured in Figure B.1 by the red line. Experiments with similar implementations (for example in the CTM) show that this will not occur at merging bottlenecks because the density will increase fast, but it can happen in situations where the capacity drops but the density only increases a little. This effect will also be evaluated when the extension is implemented in the model.

![Figure B.1: Effects of capacity drop on the shockwave derivation](image)

**B.1.3 Implementation DTM Measures**

This section proposes how traffic flows in the model can be controlled by the use of ramp metering, VSL/mainstream, control and DRG.

**Ramp metering and mainstream control**

With ramp metering and mainstream control, the sending flow of a link is controlled, which means that this measure must be implemented in equation B.5. The controller determines the maximum
flow that can leave the link \( q_{\text{measure}} \), and when the controller is active this will decrease the sending flow as expressed in equation B.11.

\[
S_i(t) = \min \left( \left( N \left( x_i^0, t + \Delta t - \frac{L_i}{V_{f,i}} \right) - N(x_i^l, t) \right), q_{M_i}, \Delta t, R_j(t), q_{\text{measure}} \Delta t \right)
\]

Depending on the objective of the traffic manager, constraints can be formulated for the controller to limit the queue length due to the measure. The most simple option is setting a minimal value for \( q_{\text{measure}} \), another option is determining a maximum queue length which is more realistic but also increases the complexity of the model.

**Dynamic Route Guidance**

The simplest way in which the route guidance can be implemented into the model is by the use of turn fractions. Since each vehicle has an origin and destination, captured by the OD matrix, these turn fractions cannot be set in general otherwise vehicles will never reach their destination. This means that for every OD combination a set of available routes must be defined, depending on the maximum detour allowed. Based on these route options and the traffic situation, vehicles can be divided over the network.

The fact that the actual turn fraction of vehicles is restricted by the routes means that the optimal turn fraction as calculated by the control method cannot always be served and the optimal situation will probably not be reached. This can be a reason not to use route guidance as instrument in the model.

**B.2 Rejection**

Although it was possible to remove the algorithm and to extend the model with the required features, the model is rejected for this research based on the arguments presented below:

- The node model, which determines the transition between one or more ingoing links and one or more outgoing links, is not functioning properly yet: when congestion occurs at the outgoing link, the addressing of spillback to the ingoing links is incorrect.
- Shockwaves and blocking back are captured in the model equations for the sending and receiving flows. These equations demand calculations of the situation at time moments outside the discrete time steps. This gives extra calculations, which increases the complexity of the model.
- The backward wave speed is calculated by the slope of the right part of the fundamental diagram (see Figure B.2), which is not correct according to the traffic flow theory.

![Figure B.2: Derivation shockwaves](image)

- It is expected that is possible to make the model piecewise affine, but also that this will take much time. The time for this research is limited, what makes the model less suitable.
C. Transformation of PAN model

Paragraph 5.2 describes the way in which the non-linear equation of the PAN model can be transformed into linear equation by adding auxiliary and binary variables. The complete transformation of equation 4.34 is given in this appendix.

The equation for the inflow of link \( x+1 \) is split up into single if/else statements:

\[
I_{x+1}(k) = \begin{cases} 
(a(k), & \text{if } (w_{x+1}(k - 1) + S_x(k)) \geq \left(\rho_{jam}L_{x+1} + O_{x+1}(k)\right) \\
(b(k), & \text{if } (w_{x+1}(k - 1) + S_x(k)) < \left(\rho_{jam}L_{x+1} + O_{x+1}(k)\right)
\end{cases}
\]

\[
a(k) = \min(\rho_{jam}L_{x+1} + O_{x+1}(k) - w_{x+1}(k) - S_x(k))
\]

\[
b(k) = \begin{cases} 
c, & \text{if } S_x(k) \leq q_{\text{cap},x+1}t_k \\
\min(S_x(k), q_{\text{dch},x+1}t_k), & \text{if } S_x(k) > q_{\text{cap},x+1}t_k
\end{cases}
\]

\[
c(k) = \begin{cases} 
S_x(k), & \text{if } w_x(k - 1) = 0 \\
\min(S_x(k), q_{\text{dch},x+1}t_k), & \text{if } w_x(k - 1) > 0
\end{cases}
\]

At first, the minimum functions will be removed using

\[
[\delta_1(k) = 1] \leftrightarrow [\rho_{jam}L_{x+1} + O_{x+1}(k) - w_{x+1}(k) - S_x(k) \geq 0]
\]

\[
[\delta_2(k) = 1] \leftrightarrow [S_x(k) - q_{\text{dch},x+1}t_k \geq 0]
\]

This gives the new equations for \( a(k), b(k) \) and \( c(k) \):

\[
a(k) = \delta_1(k) \left(\rho_{jam}L_{x+1} + O_{x+1}(k) - w_{x+1}(k - 1)\right) - \delta_1(k)S_x(k) + S_x(k)
\]

\[
b(k) = \begin{cases} 
\delta_2(k)S_x(k) - \delta_2(k)q_{\text{dch},x+1}t_k + q_{\text{dch},x+1}t_k, & \text{if } S_x(k) \leq q_{\text{cap},x+1}t_k \\
\delta_2(k)S_x(k) - \delta_2(k)q_{\text{dch},x+1}t_k + q_{\text{dch},x+1}t_k, & \text{if } S_x(k) > q_{\text{cap},x+1}t_k
\end{cases}
\]

\[
c(k) = \begin{cases} 
S_x(k), & \text{if } w_x(k - 1) = 0 \\
\delta_2(k)S_x(k) - \delta_2(k)q_{\text{dch},x+1}t_k + q_{\text{dch},x+1}t_k, & \text{if } w_x(k - 1) > 0
\end{cases}
\]

With the constraints for \( \delta_1(k) \) and \( \delta_2(k) \):

\[-m_1\delta_1(k) \leq \rho_{jam}L_{x+1} + O_{x+1}(k) - w_{x+1}(k - 1) - S_x(k) - m_1\]

\[-(M_1 - \varepsilon)\delta_1(k) \geq -\left(\rho_{jam}L_{x+1} + O_{x+1}(k) - w_{x+1}(k - 1) - S_x(k)\right) - \varepsilon\]

\[-m_2\delta_2(k) \leq S_x(k) - q_{\text{dch},x+1}t_k - m_2\]

\[-(M_2 - \varepsilon)\delta_2(k) \geq -(S_x(k) - q_{\text{dch},x+1}t_k) - \varepsilon\]

The second step is removing the if/else statements. This can be done by introducing the binary variables:

\[
[\delta_3(k) = 1] \leftrightarrow [\rho_{jam}L_{x+1} - w_{x+1}(k - 1) \geq 0]
\]

\[
[\delta_3(k) = 1] \leftrightarrow [q_{\text{cap},x+1}t_k - S_x(k) \geq 0]
\]

\[
[\delta_4(k) = 1] \leftrightarrow [w_x(k - 1) = 0]
\]

Now the equations can be transformed:

\[
I_{x+1}(k) = \delta_3(k)b - \delta_3(k)a + a
\]
For the binary variables $\delta_3(k)$, $\delta_4(k)$ and $\delta_5(k)$, the following constraints are relevant:

\[-m_3 \delta_3(k) \leq \rho_{ jam} x_{k+1} - w_{x+1}(k-1) - m_3\]
\[-(m_3 - \varepsilon) \delta_3(k) \geq -((\rho_{ jam} x_{k+1} - w_{x+1}(k-1)) - \varepsilon)\]
\[-m_4 \delta_4(k) \leq q_{cap,x+1} t_k - S_x(k) - m_4\]
\[-(m_3 - \varepsilon) \delta_4(k) \geq -(q_{cap,x+1} t_k - S_x(k)) - \varepsilon\]
\[-m_5 \delta_5(k) \leq w_x(k-1) - m_5\]
\[-(m_3 - \varepsilon) \delta_5(k) \geq -w_x(k-1) - \varepsilon\]

Now all the minimum functions and if/else statements are removed, the equation for the inflow of link $x+1$ can be filled:

\[I_{x+1}(k) = \delta_3(k) \delta_4(k) \delta_5(k) S_x(k) - \delta_3(k) \delta_4(k) \delta_5(k) \delta_2(k) S_x(k)
+ \delta_3(k) \delta_4(k) \delta_5(k) \delta_2(k) q_{dch,x+1} t_k - \delta_3(k) \delta_4(k) \delta_5(k) q_{dch,x+1} t_k
+ \delta_3(k) \delta_2(k) S_x(k) - \delta_3(k) \delta_2(k) q_{dch,x+1} t_k + \delta_3(k) q_{dch,x+1} t_k
- \delta_3(k) \delta_1(k) \rho_{ jam} x_{k+1} - \delta_3(k) \delta_1(k) S_{x+1}(k) + \delta_3(k) \delta_1(k) w_{x+1}(k-1)
+ \delta_3(k) \delta_1(k) S_x(k) - \delta_3(k) S_x(k) + \delta_1(k) \rho_{ jam} x_{k+1} + \delta_1(k) S_{k+1}(k)
- \delta_3(k) w_{x+1}(k-1) - \delta_3(k) S_x(k) + S_x(k)\]

Here, all the multiplications of binary variables must be replaced by a single binary variable:

- $\delta_3(k) \delta_4(k) = \delta_6(k)$
- $\delta_5(k) \delta_4(k) = \delta_7(k)$
- $\delta_2(k) \delta_5(k) = \delta_8(k)$
- $\delta_2(k) \delta_3(k) = \delta_9(k)$
- $\delta_1(k) \delta_3(k) = \delta_{10}(k)$

With the constraints:

\[-\delta_3(k) + \delta_6(k) \leq 0\]
\[-\delta_4(k) + \delta_6(k) \leq 0\]
\[\delta_3(k) + \delta_4(k) - \delta_6(k) \leq 1\]
\[-\delta_5(k) + \delta_7(k) \leq 0\]
\[-\delta_6(k) + \delta_7(k) \leq 0\]
\[\delta_5(k) + \delta_6(k) - \delta_7(k) \leq 1\]
\[-\delta_2(k) + \delta_8(k) \leq 0\]
\[-\delta_7(k) + \delta_8(k) \leq 0\]
\[\delta_2(k) + \delta_7(k) - \delta_8(k) \leq 1\]
\[-\delta_3(k) + \delta_9(k) \leq 0\]
\[-\delta_4(k) + \delta_9(k) \leq 0\]
\[\delta_3(k) + \delta_9(k) \leq 1\]
\[\delta_2(k) + \delta_3(k) - \delta_9(k) \leq 1\]
This gives:

\[
I_{x+1}(k) = \delta_7(k)S_x(k) - \delta_8(k)S_x(k) + \delta_9(k)q_{dch,x+1}t_k - \delta_7(k)q_{dch,x+1}t_k + \delta_9(k)S_x(k) - \delta_9(k)q_{dch,x+1}t_k + \delta_5(k)\rho_{lam}L_{x+1} - \delta_1(k)O_{x+1}(k) + \delta_1(k)w_{x+1}(k-1) + \delta_1(k)S_x(k) - \delta_4(k)S_x(k) + \delta_1(k)\rho_{lam}L_{x+1} + \delta_1(k)O_{x+1}(k) - \delta_1(k)w_{x+1}(k-1) - \delta_1(k)S_x(k) + S_x(k)
\]

Multiplications of a binary variable and a model parameter are not linear, but can be replaced by auxiliary variables:

- \(\delta_7(k)S_x(k) = z_1(k)\)
- \(\delta_8(k)S_x(k) = z_2(k)\)
- \(\delta_9(k)S_x(k) = z_3(k)\)
- \(\delta_{10}(k)S_x(k) = z_4(k)\)
- \(\delta_{10}(k)O_{x+1}(k) = z_5(k)\)
- \(\delta_{10}(k)w_{x+1}(k-1) = z_6(k)\)
- \(\delta_1(k)w_{x+1}(k-1) = z_{10}(k)\)

With the following constraints:

\[
\begin{align*}
z_1(k) & \leq M_{11}\delta_7(k) \\
z_1(k) & \geq m_{11}\delta_7(k) \\
z_1(k) & \leq S_x(k) - m_{11}(1 - \delta_7(k)) \\
z_1(k) & \geq S_x(k) - M_{11}(1 - \delta_7(k)) \\
z_2(k) & \leq M_{12}\delta_8(k) \\
z_2(k) & \geq m_{12}\delta_8(k) \\
z_2(k) & \leq S_x(k) - m_{12}(1 - \delta_8(k)) \\
z_2(k) & \geq S_x(k) - M_{12}(1 - \delta_8(k)) \\
z_3(k) & \leq M_{13}\delta_9(k) \\
z_3(k) & \geq m_{13}\delta_9(k) \\
z_3(k) & \leq S_x(k) - m_{13}(1 - \delta_9(k)) \\
z_3(k) & \geq S_x(k) - M_{13}(1 - \delta_9(k)) \\
z_4(k) & \leq M_{14}\delta_{10}(k) \\
z_4(k) & \geq m_{14}\delta_{10}(k) \\
z_4(k) & \leq S_x(k) - m_{14}(1 - \delta_{10}(k)) \\
z_4(k) & \geq S_x(k) - M_{14}(1 - \delta_{10}(k)) \\
z_5(k) & \leq M_{15}\delta_3(k) \\
z_5(k) & \geq m_{15}\delta_3(k) \\
z_5(k) & \leq S_x(k) - m_{15}(1 - \delta_3(k)) \\
z_5(k) & \geq S_x(k) - M_{15}(1 - \delta_3(k)) \\
z_6(k) & \leq M_{16}\delta_1(k)
\end{align*}
\]
\[ z_6(k) \leq m_{16} \delta_1(k) \]
\[ z_6(k) \leq S_{x}(k) - m_{16}(1 - \delta_1(k)) \]
\[ z_6(k) \leq S_{x}(k) - M_{16}(1 - \delta_1(k)) \]
\[ z_7(k) \leq M_{17} \delta_{10}(k) \]
\[ z_7(k) \leq m_{17} \delta_{10}(k) \]
\[ z_7(k) \leq o_{x+1}(k) - m_{17}(1 - \delta_{10}(k)) \]
\[ z_7(k) \leq o_{x+1}(k) - M_{17}(1 - \delta_{10}(k)) \]
\[ z_8(k) \leq M_{18} \delta_1(k) \]
\[ z_8(k) \leq m_{18} \delta_1(k) \]
\[ z_8(k) \leq o_{x+1}(k) - m_{18}(1 - \delta_1(k)) \]
\[ z_8(k) \leq o_{x+1}(k) - M_{18}(1 - \delta_1(k)) \]
\[ z_9(k) \leq M_{19} \delta_{10}(k) \]
\[ z_9(k) \leq m_{19} \delta_{10}(k) \]
\[ z_9(k) \leq w_{x+1}(k - 1) - m_{19}(1 - \delta_{10}(k)) \]
\[ z_9(k) \leq w_{x+1}(k - 1) - M_{19}(1 - \delta_{10}(k)) \]
\[ z_{10}(k) \leq M_{20} \delta_1(k) \]
\[ z_{10}(k) \leq m_{20} \delta_1(k) \]
\[ z_{10}(k) \leq w_{x+1}(k - 1) - m_{20}(1 - \delta_1(k)) \]
\[ z_{10}(k) \leq w_{x+1}(k - 1) - M_{20}(1 - \delta_1(k)) \]

That makes the expression of the inflow of link \( x + 1 \) linear:

\[ I_{x+1}(k) = z_1(k) - z_2(k) + \delta_6(k) q_{dch, x+1} t_k - \delta_7(k) q_{dch, x+1} t_k + z_3(k) - \delta_8(k) q_{dch, x+1} t_k + \delta_9(k) q_{dch, x+1} t_k - \delta_{10}(k) \rho_{jam} L_{x+1} - z_{10}(k) + z_4(k) - z_5(k) + \delta_1(k) \rho_{jam} L_{x+1} + z_8(k) - z_{10}(k) - z_6(k) + S_x(k) \]
D. CASE OPTIMISATIONS

In chapter 6, the correct functioning of the control method is verified by performing MILP optimisations for couple of cases as identified in section 6.1. This appendix will give the model equations for the network, the input for the cplex toolbox and the complete output for the first case “preventing the capacity drop”.

D.1 MODEL EQUATIONS

For the trivial network, the following linear equations express the situation in the network:

\[ w_1(k) = w_1(k-1) + I_1(k) - I_2(k) \]
\[ w_2(k) = w_2(k-1) + I_2(k) - I_3(k) \]
\[ S_1(k) = I_1(k) + w_1(k-1) \]
\[ S_2(k) = I_2(k) + w_2(k-1) \]
\[ I_1(k) = D_k t_k \]

The equations for the inflow into link 2 and link 3 are more complex. Since there is not a bottleneck located between link 1 and link 2, the inflow into link 2 is calculated by:

\[ I_2(k) = \begin{cases} 
\rho_{jam} L_2 + q_{dch,b} t_k & \text{if } w_2(k-1) + q_{c1}(k) t_k \geq (\rho_{jam} L_2 + q_{dch,b} t_k - w_2(k-1), q_{c1}(k) t_k) \\
q_{c1}(k) t_k & \text{else} 
\end{cases} \]

Removing the statements, filling in for \( \rho_{jam} L_2 = 400 \), \( q_{dch,b} t_k = \frac{4000}{3} \) and \( t_k = \frac{1}{3} \) using

\[ \delta_1(k) = 0 \Rightarrow \left( \rho_{jam} L_2 + q_{dch,b} t_k - w_2(k-1) \right) \geq (q_{c1}(k) t_k) \]
\[ \delta_2(k) = 0 \Rightarrow \left( w_2(k-1) + q_{c1}(k) t_k \right) \geq (\rho_{jam} L_2 + q_{dch,b} t_k) \]

Gives:

\[ I_2(k) = \frac{1}{3} \delta_2(k) q_{c1}(k) - \delta_2(k) \left( \frac{5200}{3} - w_2(k-1) \right) - \frac{1}{3} \delta_1(k) q_{c1}(k) + \frac{1}{3} q_{c1}(k) + \delta_1(k) \left( \frac{5200}{3} - w_2(k-1) \right) - \frac{1}{3} \delta_2(k) q_{c1}(k) + \frac{2}{3} q_{c1}(k) \]

The following variables are added:

\[ \delta_3(k) = \delta_2(k) q_{c1}(k) \]
\[ z_2(k) = \delta_3(k) w_2(k-1) \]
\[ z_3(k) = \delta_3(k) q_{c1}(k) \]
\[ z_4(k) = \delta_2(k) w_2(k-1) \]
\[ z_5(k) = \delta_2(k) q_{c1}(k) \]

This gives the final expression for \( I_2(k) \):

\[ I_2(k) = -\frac{5200}{3} \delta_3(k) + z_2(k) + \frac{1}{3} z_3(k) + \frac{5200}{3} \delta_1(k) - z_4(k) - z_5(k) + \frac{1}{3} q_{c1}(k) \]
The inflow into link 3 is not restricted by spillback, since a free outflow is assumed:

\[
I_3(k) = \begin{cases} 
S_2(k), & S_2(k) \leq q_{cap,b} \Lambda w_2(k-1) = 0 \\
\min \left( q_{dch,b} t_k, S_2(k) \right), & S_2(k) > q_{cap,b} \Lambda w_2(k-1) > 0
\end{cases}
\]

At first the minimum function is removed using \([\delta_4(k) = 1] \leftrightarrow [q_{dch,b} t_k < S_2(k)]\):

\[
\min \left( q_{dch,b} t_k, S_2(k) \right) = \delta_4(k) q_{dch,b} t_k - \delta_4(k) S_2(k) + S_2(k)
\]

Next, the equation is split up into two single if/else statements:

\[
I_3(k) = \begin{cases} 
d(k), & q_{cap,b} - S_2(k) \geq 0 \\
\delta_4(k) q_{dch,b} t_k - \delta_4(k) S_2(k) + S_2(k), & q_{cap,b} - S_2(k) < 0
\end{cases}
\]

\[
d(k) = \begin{cases} 
S_2(k), & -w_2(k-1) = 0 \\
\delta_4(k) q_{dch,b} t_k - \delta_4(k) S_2(k) + S_2(k), & -w_2(k-1) < 0
\end{cases}
\]

Introducing \([\delta_5(k) = 1] \leftrightarrow [q_{cap,b} - S_2(k) \geq 0] \text{ and } [\delta_6(k) = 1] \leftrightarrow [-w_2(k-1) = 0]\) can transform the equations into:

\[
I_3(k) = \delta_5(k) \left( -\delta_6(k) \left( \delta_4(k) q_{dch,b} t_k - \delta_4(k) S_2(k) \right) \right) + \delta_4(k) q_{dch,b} t_k - \delta_4(k) S_2(k) + S_2(k)
\]

For transformation into a linear equation, the following variables must be added:

- \(\delta_4(k)\delta_5(k) = \delta_7(k)\)
- \(\delta_6(k)\delta_5(k) = \delta_8(k)\)
- \(\delta_6(k) S_2(k) = z_6(k)\)
- \(\delta_4(k) S_2(k) = z_7(k)\)

This gives the final expression for the inflow of link 3:

\[
I_3(k) = \frac{4000}{3} \delta_6(k) + \frac{4000}{3} \delta_4(k) - z_7(k) + S_2(k)
\]

**D.2 MILP INPUT**

Software toolbox Tomlab/Cplex is used to optimise the given system \(x\), and needs the following input parts:

- Cost function which must be minimized: \(f\) (nx by 1 matrix)
- Equalities: \(Aeq x = beq\) (Aeq is ne by nx matrix, beq is ne by 1 matrix)
- Constraints: \(Ax \leq b\) (A is nc by nx matrix, b is nc by 1 matrix)
- Lower bound (lb) and upper bound (ub) of system x (both nx by 1 matrix)
- Definition which variables are integers: \(intvars\) (integer vector)

When the optimisation is completed, the following output variables are given:

- All values of the parameters of system x
- Value of the cost function: \(fval\)
D.2.1 Cost Function

In paragraph 4.4, the objective of the optimisation is defined as minimizing the total time spent (TTS) in the network and can be expressed as the total time vehicles spent in queue:

\[ J_{TTS}(S,K) = \sum_{k=0}^{K} \frac{w(k) + w(k+1)}{2} \cdot t_k \]

\( w(k) = w_1(k) + w_2(k) \): there is assumed free outflow at link 3.
\( w_x(0) = 0 \) for every \( x = 1,\ldots,3 \) since a non-congested state is assumed at the start.

With \( t_k = \frac{1}{3} \) this gives the notation used for the input of the MILP optimisation:

\[ f = \frac{1}{6} w_1(1) + \frac{1}{3} w_1(2) + \frac{1}{6} w_1(3) + \frac{1}{6} w_2(1) + \frac{1}{3} w_2(2) + \frac{1}{6} w_2(3) \]

D.2.2 Equalities

The equations of the PAN model that are identified for each \( k = 1,2,3 \) and will be used to fill the matrices of \( A_{eq} \) and \( b_{eq} \):

\[
\begin{align*}
\text{beq} & \quad A_{eq} \\
2000 & = \quad l_1(1) \\
5000/3 & = \quad l_1(2) \\
1000 & = \quad l_1(3) \\
0 & = -\frac{5200}{3} \delta_3(1) + \frac{1}{3} z_3(1) + \frac{5200}{3} \delta_1(1) - z_4(1) - z_5(1) + \frac{1}{3} q_{c1}(1) - l_2(1) \\
0 & = -\frac{5200}{3} \delta_3(2) + \frac{1}{3} z_3(2) + \frac{5200}{3} \delta_1(2) - z_4(2) - z_5(2) + \frac{1}{3} q_{c1}(2) - l_2(2) \\
0 & = -\frac{5200}{3} \delta_3(3) + \frac{1}{3} z_3(3) + \frac{5200}{3} \delta_1(3) - z_4(3) - z_5(3) + \frac{1}{3} q_{c1}(3) - l_2(3) \\
0 & = \frac{4000}{3} \delta_6(1) + \frac{1}{3} z_6(1) + \frac{4000}{3} \delta_4(1) - S_2(1) - l_3(1) \\
0 & = \frac{4000}{3} \delta_6(2) + \frac{1}{3} z_6(2) + \frac{4000}{3} \delta_4(2) - S_2(2) - l_3(2) \\
0 & = \frac{4000}{3} \delta_6(3) + \frac{1}{3} z_6(3) + \frac{4000}{3} \delta_4(3) - S_2(3) - l_3(3) \\
0 & = \quad l_1(1) - S_1(1) \\
0 & = \quad l_1(2) + w_1(1) - S_1(2) \\
0 & = \quad l_1(3) + w_1(2) - S_1(3) \\
0 & = \quad l_2(1) - S_2(1) \\
0 & = \quad l_2(2) + w_2(1) - S_2(2) \\
0 & = \quad l_2(3) + w_2(2) - S_2(3) \\
0 & = \quad l_3(1) - l_2(1) - w_1(1) \\
0 & = \quad l_3(2) - l_2(2) + w_1(1) - w_1(2) \\
0 & = \quad l_3(3) - l_2(3) + w_1(2) - w_1(3) \\
0 & = \quad l_3(1) - l_3(1) - w_2(1) \\
0 & = \quad l_3(2) - l_3(2) + w_2(1) - w_2(2) \\
0 & = \quad l_3(3) - l_3(3) + w_2(2) - w_2(3) 
\end{align*}
\]
The constraints bellowing to the 8 binary variables and the 7 auxiliary variables are given below:

\[ \begin{array}{c|c}
  b & A \\
  \hline
  2000/3 & 800/3 \delta_1(1) \\
  - 400 & -400\delta_1(1) \\
  2000/3 & w_2(1) + 800/3\delta_1(2) \\
  - 400 & -w_2(1) - 400\delta_1(2) \\
  2000/3 & w_2(2) + 800/3\delta_1(3) \\
  - 400 & -w_2(2) - 400\delta_1(3) \\
  2000 & S_2(1) + 1000/3\delta_1(1) \\
  - 5000/3 & -S_2(1) - 5000/3\delta_1(1) \\
  2000 & S_2(2) + 1000/3\delta_1(2) \\
  - 5000/3 & -S_2(2) - 5000/3\delta_1(2) \\
  2000 & S_2(3) + 1000/3\delta_1(3) \\
  - 5000/3 & -S_2(3) - 5000/3\delta_1(3) \\
  0 & -2000\delta_1(1) + z_2(1) \\
  0 & -2000\delta_1(2) + z_2(2) \\
  0 & -2000\delta_1(3) + z_2(3) \\
  0 & -2000\delta_2(1) + z_2(1) \\
  0 & -2000\delta_2(2) + z_2(2) \\
  0 & -2000\delta_2(3) + z_2(3) \\
  2000 & S_2(1) + 2000\delta_2(1) - z_2(1) \\
  0 & -2000\delta_2(1) + z_2(1) \\
  2000 & S_2(2) + 2000\delta_2(2) - z_2(2) \\
  0 & -2000\delta_2(2) + z_2(2) \\
  2000 & S_2(3) + 2000\delta_2(3) - z_2(3) \\
  0 & -2000\delta_2(3) + z_2(3) \\
  0 & -1/3 q_{ctrl}(1) + z_1(1) \\
  2000 & 1/3 q_{ctrl}(1) - z_1(1) + 2000\delta_1(1) \\
  0 & -2000\delta_1(1) + z_1(1) \\
  0 & -1/3 q_{ctrl}(2) + z_1(2) \\
  2000 & 1/3 q_{ctrl}(2) - z_1(2) + 2000\delta_1(2) \\
  0 & -2000\delta_1(2) + z_1(2) \\
  0 & -1/3 q_{ctrl}(3) + z_1(3) \\
  2000 & 1/3 q_{ctrl}(3) - z_1(3) + 2000\delta_1(3) \\
  0 & -2000\delta_1(3) + z_1(3)
\end{array} \]

Besides the constraints set to the logical variables \( \delta_i \in \{0,1\} \) and auxiliary real variables \( z_i \), constraints are set to the values of the controller \( q_{c1}(k) \): the output of the controller at the end of link 1 may never exceed the sending flow \( S_1(k) \). This is captured in the following constraints:

\[
\begin{align*}
q_{c1}(1) & \leq S_1(1) \\
q_{c1}(2) & \leq S_1(2) \\
q_{c1}(3) & \leq S_1(3)
\end{align*}
\]
D.2.4 **Boundaries**

Lower and upper boundaries must be set for each parameter in the PAN-model.

In this case the lower boundaries of all system elements are 0, except the controller parameters to ensure a minimum outflow of the controller.

\[
\begin{align*}
\text{lb} \\
q_{\text{ctrl}}(1) & \geq 1000 \\
q_{\text{ctrl}}(2) & \geq 1000 \\
q_{\text{ctrl}}(3) & \geq 1000
\end{align*}
\]

The upper boundaries indicate the maximum value of the parameters. The upper bound of all the binary variables \( \delta_i(k) \) is 1.

\[
\begin{align*}
\text{ub} \\
q_{\text{ctrl}}(k) & \leq 6000 \\
I_1(k) & \leq D_1 t_k \\
I_2(k) & \leq 2000 \\
I_3(k) & \leq 2000 \\
S_1(k) & \leq 2000 \\
S_2(k) & \leq 2000 \\
w_1(k) & \leq 600 \\
w_2(k) & \leq 800 \\
z_3(k) & \leq 2000 \\
z_4(k) & \leq 2000
\end{align*}
\]

D.2.5 **Integers**

All the binary variables \( \delta_i(k) \) is indicated by a vector of length \( x \). When a variable of system \( x \) is a binary variable, 1 is filled in, otherwise 0.

D.3 **MILP Output**

All the matrices and vectors that are identified in the previous section are implemented into MatLab and the optimisation is run by TOMLAB Cplex. This gives the optimal solution for the cost function: the total time spent in queue is about 167 hours. It also return all the values of the parameters of system \( x \), which are presented in Table D.1. The first three values represent the optimal settings for the controllers (5000,5000 and 4000), that could be stated before since the capacity of the bottleneck is 5000 vehicles per hour and the maximal outflow of link 3 can therefore never exceed 5000.

The output of the MILP optimisation is also checked by comparison with the output of the non-linear model whit the same settings for the controller. This gives exactly the same output for the parameters and the objective function.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Output MILP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_{ctrl}(1)$</td>
<td>5.000</td>
</tr>
<tr>
<td>$q_{ctrl}(2)$</td>
<td>5.000</td>
</tr>
<tr>
<td>$q_{ctrl}(3)$</td>
<td>4.000</td>
</tr>
<tr>
<td>$w_1(1)$</td>
<td>333.3</td>
</tr>
<tr>
<td>$w_1(2)$</td>
<td>333.3</td>
</tr>
<tr>
<td>$w_1(3)$</td>
<td>0</td>
</tr>
<tr>
<td>$w_2(1)$</td>
<td>0</td>
</tr>
<tr>
<td>$w_2(2)$</td>
<td>0</td>
</tr>
<tr>
<td>$w_2(3)$</td>
<td>0</td>
</tr>
<tr>
<td>$I_1(1)$</td>
<td>2.000</td>
</tr>
<tr>
<td>$I_1(2)$</td>
<td>1.667</td>
</tr>
<tr>
<td>$I_1(3)$</td>
<td>1.000</td>
</tr>
<tr>
<td>$I_2(1)$</td>
<td>1.667</td>
</tr>
<tr>
<td>$I_2(2)$</td>
<td>1.667</td>
</tr>
<tr>
<td>$I_2(3)$</td>
<td>1.333</td>
</tr>
<tr>
<td>$I_3(1)$</td>
<td>1.667</td>
</tr>
<tr>
<td>$I_3(2)$</td>
<td>1.667</td>
</tr>
<tr>
<td>$I_3(3)$</td>
<td>1.333</td>
</tr>
<tr>
<td>$S_1(1)$</td>
<td>2.000</td>
</tr>
<tr>
<td>$S_1(2)$</td>
<td>2.000</td>
</tr>
<tr>
<td>$S_1(3)$</td>
<td>1.333</td>
</tr>
<tr>
<td>$S_2(1)$</td>
<td>1.667</td>
</tr>
<tr>
<td>$S_2(2)$</td>
<td>1.667</td>
</tr>
<tr>
<td>$S_2(3)$</td>
<td>1.333</td>
</tr>
<tr>
<td>$\delta_1(1)$</td>
<td>1</td>
</tr>
<tr>
<td>$\delta_1(2)$</td>
<td>1</td>
</tr>
<tr>
<td>$\delta_1(3)$</td>
<td>1</td>
</tr>
<tr>
<td>$\delta_2(1)$</td>
<td>1</td>
</tr>
<tr>
<td>$\delta_2(2)$</td>
<td>1</td>
</tr>
<tr>
<td>$\delta_2(3)$</td>
<td>1</td>
</tr>
<tr>
<td>$z_1(1)$</td>
<td>1.667</td>
</tr>
<tr>
<td>$z_1(2)$</td>
<td>1.667</td>
</tr>
<tr>
<td>$z_1(3)$</td>
<td>1.333</td>
</tr>
<tr>
<td>$z_1(1)$</td>
<td>1.667</td>
</tr>
<tr>
<td>$z_2(2)$</td>
<td>1.667</td>
</tr>
<tr>
<td>$z_2(3)$</td>
<td>1.333</td>
</tr>
</tbody>
</table>

*Table D.1: Output MILP optimisation*