Static Traffic Assignment with Queuing

Master Thesis Applied Mathematics

A.S. van Leeuwen

March 25, 2011

UNIVERSITY OF TWENTE.

Goudappel Coffeng
Consultants on traffic and transport

Static Traffic Assignment with Queuing

A.S. van Leeuwen

0067857

University of Twente

Applied Mathematics

Industrial Engineering and Operations Research

Discrete Mathematics and Mathematical Programming

Graduation committee:

Dr. G.J. Still (University of Twente)

Dr. W. Kern (University of Twente)

Prof. Dr. M.J. Uetz (University of Twente)

Ir. L.J.N. Brederode (Goudappel Coffeng BV)

Abstract

Traffic models are used to predict traffic flows and travel times on a road network. Due to the ever increasing level of congestion at highways and urban areas, the need for better results from those models also increases. On the other hand, these predictions should be available in a reasonable time such that the right action to avoid congestion can be taken quickly. Because of computation time issues, often static models instead of dynamic models are used.

For realistic predictions a traffic model should be able to take capacity constraints into account, something that traditional static traffic assignment models cannot do directly. Besides that, queuing and spillback phenomena should also be taken care of. In this thesis, an overview is given of the static models and concepts in literature that in some way or another do account for capacity limitations and/or spillback effects. It is explained that all of these models still have drawbacks, and therefore a new static traffic assignment model is introduced.

This new model is basically obtained in two successive steps. First the static variant of the dynamic network loading model LTM (Link Transmission Model) by Yperman (2007) is derived by assuming stationary traffic demand. After that, the STAQ (Static Traffic Assignment with Queuing) model is created by assuming instantaneous traffic flow propagation.

In this model queues are constructed according to kinematic wave theory using the fundamental diagram, which gives the model a mathematical basis. Also spillback effects are accounted for. Therefore, the model is in theory superior to other static traffic assignment models, but empirical tests are needed to provide hard evidence. STAQ is currently in development at Goudappel Coffeng B.V.

Preface

This thesis is the result of my research at Goudappel Coffeng B.V. and the University of Twente on static traffic assignment models. I worked for about 10 months on this research, which was done at the chair of Discrete Mathematics and Mathematical Programming.

I would like to thank my supervisors at Goudappel in Deventer, Luuk and Michiel, for the daily guidance during the research. They were always willing to discuss problems and gave many suggestions for improvements.

Also I would like to thank Walter and Georg, my supervisors at the university. Though sometimes I had difficulties explaining the idea and the workings of the model, they were very helpful with improving my report.

Finally I would like to thank my friends and family, especially my parents, for their patience and support, not in the least place financially.

Contents

1. Introduction

Traffic models are widely used as a tool to assist in making decisions in mobility and infrastructure planning. These tools can help in analyzing congestion issues, and can predict the consequences of certain measurements, like adding an extra lane, building a new road, raising the maximum allowed speed or applying road pricing.

A classical traffic model to predict traffic flows is known as the four step model. The four steps are trip generation, trip distribution, modal split and assignment. In this model the network is divided in zones, between which people can travel. The first step is the trip generation. Based on data about the population and economic activity like employment, shopping space, educational and recreational facilities at a zone, the total number of trips generated at each zone and the number of trips attracted by each zone is estimated. At the next step, these trips are distributed over space, resulting in a OD-matrix (origin-destination matrix). An OD-matrix contains the number of trips from each origin to each destination (zone). The third step is called modal split, in which the trips are allocated to different modes of transport, for instance car, public transport, or bike. In the last step, the assignment of the trips to the network links is done. Basically this means that for each OD-pair, the number of trips is distributed over a number of routes (or paths) which consists of a number of links in sequence. This gives the number of vehicles that want to travel over each link, i.e. the demand for each link.

Traffic models can be either static or dynamic. Within dynamic models, the traffic demand (OD-matrix) can change through time, while within a static model it is assumed that the traffic demand is constant through time.

In this thesis, the focus will be on the assignment phase of the process of a static traffic model. As the mode of transport, only car is considered. It is assumed that an OD-matrix is available.

The assignment phase of the traffic model consists of two components (see Figure 1). The first component is the route choice model, in which routes are selected and the traffic demand is assigned to these routes. The second component is the network loading model, which takes the assigned route flow as input and describes the way in which the traffic is propagated through the network. The network loading model yields the link flows which can be used to determine the route travel times.

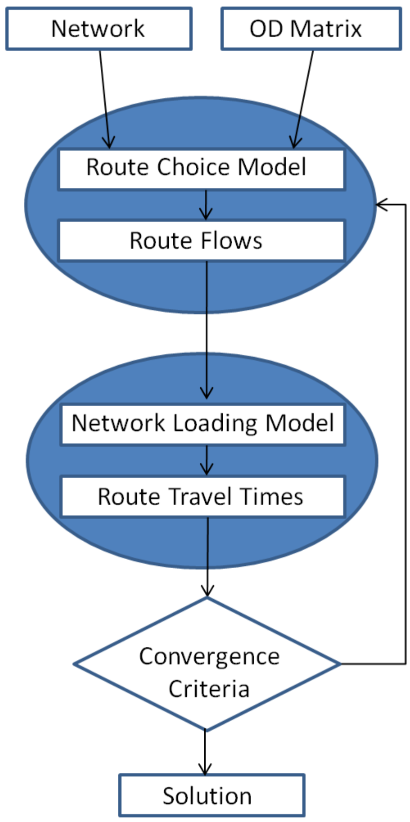


Figure 1. Schematic overview of the traffic assignment model.

1.1. Motivation of research

With the ever increasing level of congestion it becomes more and more important that the traffic model predictions are as accurate as possible, but on the other hand the computation time should be as low as possible such that these predictions can be achieved very fast. Dynamic models can produce realistic predictions with consistent congestion formation and delays, but in general take much more time to compute than static models. On the other hand, current static models can compute a result very fast, but these models usually cannot handle congestion effects correctly as these effects are time dependent in nature. Therefore, there is need for a quasi-dynamic model, that produces realistic travel times but is still computational attractive.

1.2. Research questions

In this thesis it will be investigated whether a traffic assignment model can be formulated that can produce realistic travel times in a static or quasi-dynamic context and can generate these results in a reasonable time. To accomplish this objective, the following research questions are formulated. At first in the existing literature will be searched for a model that complies with the demands mentioned before:

1. Are there static traffic assignment models in the literature that can deal with congestion and can compute realistic travel times?

If such a model cannot be found, the goal is to specify such a model:

2. Can a model be specified that propagates traffic correctly through the network and computes more realistic travel times than current static models, and can still compute them in a reasonable time?

1.3. Overview of the thesis

In Figure 2 a schematic overview of the thesis is given. In the next chapter, the first research question is answered. A summary is given of the different static traffic assignment models that can deal with capacity constraints. The different approaches are discussed and the choice to derive a new model from the dynamic Link Transmission Model (LTM) by Yperman (2007) is motivated.

In chapter 3, the working of the LTM by Yperman is presented. It is explained how the LTM is used as the basis of the new model. After that the quasi-dynamic variant of the LTM with stationary traffic demand is derived.

The subject of the fourth chapter is STAQ: Static Traffic Assignment with Queuing, the newly developed model at Goudappel Coffeng B.V. The model uses elements from the LTM by Yperman, with some additional assumptions. The most important additional assumption is that of instantaneous travel flow propagation. The derivation of STAQ from the LTM with stationary demand is showed and an algorithm is presented which can solve the STAQ model efficiently.

Finally some conclusions and recommendations for further work are given in the last chapter.

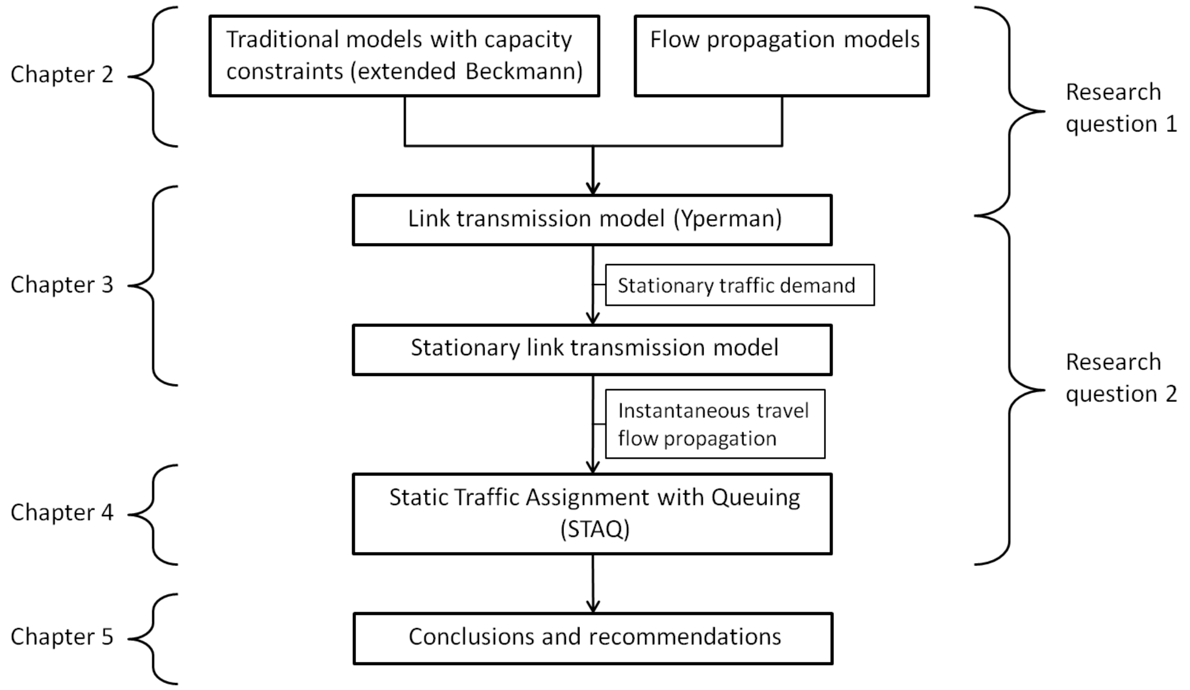


Figure 2. Research approach and structure of the thesis.

2. Traffic assignment models with capacity constraints: A summary

In this chapter an overview is given of different traffic models that somehow can deal with capacity constraints. All the models considered here provide a method to predict the traffic flow on the links, given a network and an OD-matrix with traffic demand. From this information the travel times can be determined.

The traffic demand needs to be assigned to the links in the network in a certain way. To obtain realistic results, the capacities of the links in the network and the congestion that it can cause must be taken into account. In this overview different approaches are discussed that are proposed in the literature. In the first part, the traditional approach is explained, that does not take the capacity constraints into account. After that, the traditional approach extended with capacity constraints is discussed, with a number of different solution techniques. Then some other, more realistic ideas are presented. Finally the advantages and disadvantages of the models are discussed and it is explained what qualities the new model should have.

2.1. Traditional static traffic assignment

Wardrop's first principle (Wardrop 1952) states that at an equilibrium situation the travel times of all routes between each OD-pair actually used are the same, and there are no unused routes with a lower travel time. This means that no user can lower his travel time by switching individually to another route. To find an equilibrium solution in a network, the problem can be described in a mathematical way (Beckmann e.a. 1956).

Consider a network (V, E) given by a set of nodes (or vertices) V and a set of links (or edges) E. The set W consists of the origin-destination (OD) pairs $(s_w, t_w) \in V \times V$, with a demand $d_w \geq 0$, $w \in W$ for each OD-pair. Furthermore, $\mathcal{P}_w$ is the set of directed (s_w, t_w) routes (or paths) for OD-pair w with $\mathcal{P} = \bigcup_w \mathcal{P}_w$. The objective is to find a traffic flow through the network, that satisfies the demand $d \in \mathbb{R}^W$ for each OD-pair, by assigning a certain amount of nonnegative flow x_e to each link e. More formally, find an $x \in \mathbb{R}_+^E$ satisfying:

$$\begin{aligned} & x = \Delta f \\ & f \geq 0 \\ & \Lambda f = d \end{aligned} \qquad (1)$$

where $f \in \mathbb{R}^{\mathcal{P}}$ consists of the path flows f_p for each path $p \in \mathcal{P}$, and the incidence matrices $\Delta \in \mathbb{R}^{E \times \mathcal{P}}$ and $\Lambda \in \mathbb{R}^{W \times \mathcal{P}}$ are defined as follows:

$$\Delta_{ep} = \begin{cases} 1 & if\ e\ is\ a\ link\ of\ path\ p \\ 0 & otherwise \end{cases}$$

$$\Lambda_{wp} = \begin{cases} 1 & if\ p\ is\ a\ path\ connecting\ OD\ pair\ (s_w, t_w)\ with\ p \in \mathcal{P}_w \\ 0 & otherwise \end{cases}$$

The set of feasible flows X_d consists of all traffic flows $x \in \mathbb{R}_+^E$ (with a corresponding f) such that (1) is satisfied for demand $d \in \mathbb{R}_+^W$.
For each link is usually defined a link cost function or travel time function $c_e(x_e)$ (separable) or more general $c_e(x)$ (non-separable cost function). In the separable case, the link cost is only dependent of the flow on the link itself. Usually it is assumed that the link cost functions are continuous and monotonically non-decreasing in the link flows. So, more flow means higher link costs. The function named after the United States Bureau of Public Roads (BPR function) is frequently used as a cost function:

$$c_e(x_e) = \overline{t_e}\left(1 + \alpha \frac{{x_e}^{\beta}}{u_e}\right)$$

where $\overline{t_e}$ is the free flow travel time on link e, u_e is the capacity of link e in vehicles per time unit and α and β are adjustable parameters (US Bureau of Public Roads 1964).

A common assumption is that the path costs are additive:

$$c_p(x) = \sum_{e \in p} c_e(x) \qquad \forall p \in \mathcal{P}$$

A given feasible traffic flow $x^* = \Delta f^*$ is called a Wardrop Equilibrium flow if the following holds:

$$f_p{}^* > 0 \Rightarrow c_p(x^*) \leq c_q(x^*) \qquad \forall p, q \in \mathcal{P}_w, w \in W$$

Or, equivalently:

$$f_p{}^* > 0 \Rightarrow c_p(x^*) = min_{q \in \mathcal{P}_w} c_q(x^*) \qquad \forall p \in \mathcal{P}_w, w \in W$$

So, when a given feasible flow assigns a positive flow on a certain path, the costs on that path cannot be greater than the costs on any other path for the corresponding OD-pair. In other words, if a cheaper alternative exists for a certain path, that path will have zero flow assigned:

$$c_p(x^*) > c_q(x^*) \Rightarrow f_p{}^* = 0 \qquad \forall p, q \in \mathcal{P}_w, w \in W$$

If the cost functions are separable, the equilibrium solution can be found by solving the following Beckmann User Equilibrium optimization problem (Beckmann e.a. 1956). This formulation is in general not possible if the cost functions are not separable.

$$\min_x Z(x) = \sum_{e \in E} \int_0^{x_e} c_e(s) ds \qquad (2)$$

subject to (1)

Because the set X_d is compact (closed and bounded), the (continuous) objective function will reach its minimum, so the existence of a solution is guaranteed. If all cost functions are continuous and non-decreasing, this optimization problem is convex and then it can be showed that a solution $x^* \in X_d$ is a Wardrop Equilibrium if and only if x^* solves the program (2). If the cost functions are also strictly increasing, then the link flow solution is unique (Smith 1979). However, the problem is not (strictly) convex in the path flows, so there may be more than one path flow pattern that generates the unique link flow solution. (Yang & Huang 2005)

The necessary and sufficient first order optimality conditions (Karush Kuhn Tucker conditions) for an optimal path flow vector f^* are as follows (Yang & Huang 2005):

$$f_p{}^*\left(c_p(x^*) - \mu_w\right) = 0 \qquad \forall p \in \mathcal{P}_w, w \in W$$

$$c_p(x^*) - \mu_w \geq 0 \qquad \forall p \in \mathcal{P}_w, w \in W$$

$$x^* - \Delta f^* = 0$$

$$\Lambda f^* - d = 0$$

$$f^* \geq 0$$

Here $\mu \in \mathbb{R}^W$ is the vector of Lagrange multipliers for the demand constraints corresponding to the OD pairs. It follows that for all $p \in \mathcal{P}_w, w \in W$

$$c_p(x^*) = \mu_w \qquad \text{if } f_p{}^* > 0$$

and

$$c_p(x^*) \geq \mu_w \qquad \text{if } f_p{}^* = 0$$

This corresponds to the definition of a Wardrop Equilibrium. μ_w can be interpreted as the equilibrium (minimal) path cost for OD pair w.

When the cost functions are non-separable, in general the problem cannot be written in the Beckmann formulation. However, it can be formulated as a Variational Inequality Problem (VIP):

$$\begin{aligned} &\text{find} && x^* \in X_d \\ &\text{such that} && c(x^*)^{\mathrm{T}}(x - x^*) \geq 0 \qquad \forall x \in X_d \end{aligned}$$

where $c(x) \in \mathbb{R}^E$ is the vector of link costs for each link for link flows x.
If the link cost functions only depend on the flow on the link itself (separable cost functions), the VIP is equivalent to the Beckmann minimization problem.

The standard Beckmann model can be solved for instance with a Frank-Wolfe type algorithm. This method converts the problem into a series of shortest path subproblems which can be solved relatively easy. Most solution methods exploit the Cartesian product structure of the feasible solution set. The idea is as follows:

1. (Start) Perform an all-or-nothing assignment to obtain link flows based on the current link travel times. All-or-nothing assignment means that each OD-pair only uses the shortest route for all its demand.

2. (Iteration) Find a feasible search direction by finding the shortest paths.

3. Find the optimal step size (line search) in this direction.

4. Update the link flows.

5. If the convergence criterion is not met, continue at step 2.

2.2. Traditional static traffic assignment with capacity constraints

The classical Beckmann model as described in the previous section does not take link capacities explicitly into account and therefore can produce unrealistic flows. Assigning more flow to a link than its capacity is penalized by the monotonically non-decreasing travel time function, but it is not forbidden. Numerous ideas and suggestions have been given in the literature to deal with link capacities. One can use travel time functions with vertical asymptotes at the capacity. However, this method is not favorable because it can produce unrealistically high travel times when flow approaches capacity and can give numerical problems (Larsson & Patriksson 1995). Therefore in this overview the focus is on models that have explicit link capacity constraints.
By adding simple capacity constraints to the Beckmann optimization problem, the capacitated (or extended) Beckmann User Equilibrium optimization problem is obtained. (Larsson & Patriksson 1995, Nie e.a. 2004)

$$\min_x Z(x) = \sum_{e \in E} \int_0^{x_e} c_e(s) ds \tag{3}$$

subject to (1) and

$$x_e \leq u_e \qquad \forall e \in E$$

The necessary and sufficient first order optimality conditions (Karush Kuhn Tucker conditions) for an optimal path flow vector f^* are as follows (Yang & Huang 2005):

$$f_p{}^*\left(\hat{c}_p(x^*) - \mu_w\right) = 0 \qquad \forall p \in \mathcal{P}_w, w \in W$$

$$\hat{c}_p(x^*) - \mu_w \geq 0 \qquad \forall p \in \mathcal{P}_w, w \in W$$

$$x^* - \Delta f^* = 0$$

$$\Lambda f^* - d = 0$$

$$f^* \geq 0$$

$$x_e{}^* \leq u_e \qquad \forall e \in E$$

$$\lambda_e(x_e{}^* - u_e) = 0 \qquad \forall e \in E$$

$$\lambda \geq 0$$

Here

$$\hat{c}_p(x^*) = \sum_{e \in p} \hat{c}_e({x_e}^*) \qquad \forall p \in \mathcal{P}_w, w \in W$$

are called the generalized path costs, and

$$\hat{c}_e({x_e}^*) = c_e({x_e}^*) + \lambda_e \qquad \forall e \in E$$

are the generalized link costs. The vector $\mu \in \mathbb{R}^W$ is the vector of Lagrange multipliers for the demand constraints corresponding with the OD pairs, which can be interpreted as the vector of equilibrium generalized path travel times. Similar to the standard Beckmann model, it holds for all $p \in \mathcal{P}_w, w \in W$ that

$$\hat{c}_p(x^*) = \mu_w \qquad \text{if } {f_p}^* > 0$$

and

$$\hat{c}_p(x^*) \geq \mu_w \qquad \text{if } {f_p}^* = 0$$

Furthermore, $\lambda \in \mathbb{R}^E$ is the vector of Lagrange multipliers for the capacity constraints of each link. λ can be seen as the equilibrium queuing delay time, an additional time penalty that users traveling on this link are willing to wait for continuously using this link (Yang & Huang 2005). This delay time can only be positive when the link flow equals the link capacity. While there is a unique solution for the vector of equilibrium generalized path travel times μ, the vector of delay times λ is in general not unique. Therefore, not too much importance should be given to these delay times (Nie e.a. 2004).
The capacitated Beckmann optimization problem is not capable to handle spillback, since there are no physical queues created. Spillback (or blocking back) is the phenomenon that a queue will continue on the preceding links when it reaches the link end. Within the capacitated Beckmann problem, only a time penalty is given to the links that are 'congested', i.e. to the links that have an assigned traffic flow that is equal to the capacity.

When applying the Frank Wolfe method to the capacitated Beckmann problem, the direction search subproblem changes to a multi-commodity flow problem with inequality constraints instead of a shortest path problem. This is much harder to solve (Larsson & Patriksson 1995). Therefore, different solution methods have been proposed in the literature. Most methods convert the capacitated problem into a series of uncapacitated problems, such that efficient methods for the Beckmann problem can be applied. Some methods apply an interior penalty function to penalize the usage of a congested link. Other methods use an augmented Lagrangean approach combined with an exterior penalty function. The interior penalty function adds the penalty directly to the Beckmann objective function instead of indirectly using augmented Lagrange multipliers. Another possibility is applying a dynamic penalty function. In the following sections different solution techniques are discussed.

2.2.1. Interior penalty function approach

The interior penalty function (IPF) approach tries to approximate the capacity constrained traffic assignment problem by adding a penalty term to the objective function of the "unconstrained" problem. By imposing an asymptotic penalty term when the link flow approaches the link capacity, it is prevented that an infeasible solution is achieved. An initial feasible solution is necessary for this approach. Within optimization theory, this approach is known as the barrier method (see e.g. Faigle e.a. 2002). The optimization problem becomes:

$$\min_{x \in X_d} P_c(x) = Z(x) + c \cdot \sum_{e \in E} p_e(x_e)$$

Here $c \geq 0$ is the penalty parameter, $p_e(x_e)$ is a penalty function that is continuous on $[0, u_e)$ for all $e \in E$ and

$$p_e(x_e) \to +\infty \text{ when } x_e \to u_e$$

In principle, the problem can now be solved as a standard Beckmann program. The sequence of solutions will converge to the optimal solution of (3) as $c \to 0$.

Examples of penalty functions in the IPF approach:

$$p_e(x_e) = \int_0^{x_e} \frac{1}{u_e - s} ds$$ Nie e.a. (2004)

$$p_e(x_e) = -\ln\left(\frac{u_e - x_e}{u_e}\right)$$ Prashker & Toledo (2004)

Note that the penalty functions of Nie e.a. and Prashker & Toledo are in fact identical:

$$p_e(x_e) = \int_0^{x_e} \frac{1}{u_e - s} ds = -\ln(u_e - s)]_0^{x_e}$$

$$= -\ln(u_e - x_e) - -\ln(u_e - 0) = \ln(u_e) - \ln(u_e - x_e)$$

$$= -\ln\left(\frac{(u_e - x_e)}{u_e}\right)$$

Nie e.a. (2004) solve the subproblems with a gradient projection algorithm that updates the path flows in each iteration and searches in a direction orthogonal to the previous direction. Prashker & Toledo (2004) provide a similar implementation of the interior penalty function approach. They use a path based adaptation of the gradient projection algorithm to solve the subproblems.

2.2.2. Augmented Lagrangean dual with exterior penalty function approach

In this approach, also an extra term is added to the objective function. By imposing a penalty term when the flow of a link is greater than the link capacity, the method forces the sequence of solutions into the feasible area. An initial feasible solution is not necessary. The optimization problem becomes, similar to the interior penalty function approach:

$$\min_{x \in X_d} P_c(x) = Z(x) + c \cdot \sum_{e \in E} p_e(x_e) \tag{4}$$

where $p_e(x_e)$ is a penalty function, with $p_e(x_e) \geq 0$ for all $x \in X_d$ and $p_e(x_e) = 0$ if and only if $x_e \leq u_e$, and $p_e(x_e)$ is continuous on X_d. Now the penalty subproblem is an uncapacitated traffic assignment problem again. Let the solution of (4) for penalty parameter c be $x(c)$. It can be showed that (under certain assumptions) $P_c(x(c)) \leq Z(x^*)$ for all $c > 0$, and $\lim_{c \to \infty} x(c) = x^*$, where x^* is the optimal solution of the capacitated Beckmann problem (3) (Larsson & Patriksson 1995). So, the sequence of solutions will converge to the optimal solution of (3) as $c \to \infty$. An example of a penalty function is:

$$p_e(x_e) = \tfrac{1}{2} \max{}^2(0; x_e - u_e) \qquad \forall e \in E$$

Because of the condition $c \to \infty$ the problem becomes ill-conditioned. This is inherent in the penalty approach. To avoid this, a Lagrangean term is added to the objective function, creating an augmented Lagrangean function (for the penalty function above): (Larsson & Patriksson 1995)

$$\min_{x \in X_d} L_c(x, \lambda) = Z(x) + \sum_{e \in E} \frac{1}{2c} \left(\max{}^2\{0; \lambda_e + c \cdot (x_e - u_e)\} - {\lambda_e}^2 \right) \tag{5}$$

Let $x(c, \lambda)$ be the solution of the subproblem (5) with penalty parameter c and vector λ of Lagrangean dual variables for the capacity constraints. It can be showed that for $c > 0$ large enough: there exists an optimal Lagrange multiplier λ^* such that $L_c(x(c, \lambda), \lambda) \leq L_c(x(c, \lambda), \lambda^*) = Z(x^*) \;\; \forall \lambda \geq 0$, and $\lim_{\lambda \to \lambda^*} x(c, \lambda) = x(c, \lambda^*) = x^*$.
Compare (part of) the KKT conditions of the capacitated Beckmann problem (3)

$$\nabla Z(x) + \sum_{e \in E} \lambda_e \nabla (x_e - u_e) = \nabla Z(x) + \lambda$$

and the augmented Lagrangean (5):

$$\nabla Z(x) + \nabla \sum_{e \in E} \frac{1}{2c} \left(\max{}^2\{0; \lambda_e + c \cdot (x_e - u_e)\} - {\lambda_e}^2 \right) = \nabla Z(x) + \lambda_{max}$$

with $\lambda_{max} \in \mathbb{R}^E$ is a vector with elements $\lambda_{max,e} = \max\{0; \lambda_e + c \cdot (x_e - u_e)\}$ for each $e \in E$. It follows that the sequence of Lagrange multipliers should be updated as follows:

$${\lambda_e}^{n+1} = \max\{0; {\lambda_e}^n + c^n \cdot ({x_e}^n - u_e)\}$$

It follows that if $\lambda^n \to \lambda^*$ then the condition $c^n \to \infty$ is no longer needed for convergence, so the problem is no longer ill conditioned.

2.2.3. Dynamic penalty function approach

In the method of Shahpar e.a. (2008), the capacity constraints are taken into consideration by implicitly adding a penalty function to the link travel time functions, which they call a dynamic penalty function. They let a penalty term

$$\sum_j p_j(x, \alpha_j) \frac{\partial g_j(x)}{\partial x_e}$$

play the role of the Lagrangean multipliers for the capacity constraints in the KKT conditions (based on Larsson & Patriksson (1999)). Here $g_j(x) \leq 1$ is the j^{th} side constraint and $p_j(x, \alpha_j)$ is a interior penalty function. For general side constraints the generalized path costs now become:

$$\hat{c}_p = \sum_{e \in p} \left(c(x_e) + \sum_j p_j(x, \alpha_j) \frac{\partial g_j(x)}{\partial x_e} \right); \quad \forall p \in \mathcal{P}_w, w \in W$$

The simple capacity constraints $x_e \leq u_e$ are reformulated as $g_e(x) = \frac{x_e}{u_e} \leq 1$ for all $e \in E$.
For details see Shahpar e.a. (2008). They tested their algorithm on some small and medium sized networks for the simple capacity side constraints. Their experiments achieved a solution for the link capacity constrained problem faster than the interior penalty function method or the augmented Lagrangean method.

2.2.4. Conclusion

All the above described solution techniques can solve the capacitated Beckmann optimization problem. Though the capacitated Beckmann problem has a nice mathematical formulation, in general it cannot give realistic results, since queuing and spillback effects are not taken into account. It is known from practice that congestion on a link can have consequences for traffic flow on upstream links. Besides that, the travel time functions used by these models are difficult to determine and need time for calibration. Therefore, in the next section other approaches that avoid the use of these functions are discussed.

2.3. More realistic static traffic assignment

In the following, models that do not directly use travel time functions are examined. First the stable dynamics model by Nesterov & De Palma is described. After that a number of flow propagation models (simulation models) are discussed. In these type of models, no route choice is modeled. It is assumed that a certain route flow assignment is given, and this traffic flow is propagated through the network in a certain way, taking congestion effects into account.

2.3.1. Stable equilibrium model

Nesterov & De Palma (e.g. 2000, 2003) introduce a theory of static equilibria in congested networks, the stable equilibrium model. They observe that if the flow on a link is small, then either little traffic is using it (low travel time), or the link is heavily congested (high travel time). They state that the assumption that the travel time is an increasing function of the flow as is done within the extended Beckmann formulation is artificial. Nesterov & De Palma use the fundamental relation between the flow rate (intensity), travel time and the loading (density) on a link: flow = speed x density (see also section 3.2 about the fundamental diagram).
They obtain stable equilibrium solutions only by imposing lower bounds on the travel time (the free flow travel time) and upper bounds on the flow (the link capacity). They assume that if the flow on a link is strictly smaller than the capacity, then there is no congestion and the equilibrium travel time on this link is equal to the minimal (free flow) travel time. If the link flow is equal to the capacity, then the equilibrium travel time is greater or equal then the minimal travel time. In this case there is congestion on the link. In fact this could be defined as a travel time function as follows:

$$c(x_e) = \begin{cases} \overline{t_e} & if\ 0 \leq x_e < u_e \\ M & if\ x_e = u_e \end{cases}$$

Here $\overline{t_e}$ is the free flow travel time on link e and $M \geq \overline{t_e}$.
A Nesterov Equilibrium solution (x^*, t^*) satisfies the following conditions: $x^* \leq u$ and $t_e^* \geq c(x_e^*)$, and x^* is a Wardrop Equilibrium with respect to t^*. Kern & Still (2009) show that the idea of Nesterov & De Palma can be seen as a special case of a generalized Wardrop Equilibrium, and that it can be extended to the non-separable case.
The model of Nesterov & de Palma is not capable of handling spillback. They propose to add an extra condition such that the length of the queue on a link cannot exceed the link length by stating that $t_e < L_e/l \cdot u_e \quad \forall e \in E$, where L_e is the link length and l the average length occupied by a vehicle. However, in the rest of their papers they do not use this condition; they just assume that spillback never occurs, so it is unclear what the effect is of this condition in the model.
Chudak e.a. (2007) compare the Nesterov & De Palma model with the (extended) Beckmann model on some small networks. However, it is not clear which of the models better predicts the real traffic flow. Real data from traffic counters is needed to be able to draw hard conclusions..

2.3.2. Flow propagation models

In this section some flow propagation models are described. Four models are considered: the model by Bifulco and Crisalli (1998), the model by Bundschuh e.a. (2006), QBLOK by 4Cast (2009) and the Link Transmission Model by Yperman (2007).

Bifulco and Crisalli (1998) present a model for a stochastic user equilibrium assignment problem. At first they describe a flow propagation model which takes link capacities explicitly into account. Next they describe how their model can be used within a traffic assignment model to find an equilibrium flow. The flow propagation model cannot handle spillback. They simply assume that the queue length on a link cannot grow bigger than the length of the link, thus assuming that spillback never occurs.

They assume that for a certain OD-pair, each path has a certain probability to be taken by a user, depending on the path cost vector of that OD-pair. In their model, they determine iteratively which quantity of users on a link can proceed to the next link on their path, by checking the link capacities. This results in a new link flow pattern that is used in the next iteration. The algorithm stops when a stable link flow solution is established, which is close enough to the previous iteration. More precisely they state:

$$x = \Psi P d$$

Where the (e, p)-element of $\Psi \in \mathbb{R}^{E \times \mathcal{P}}$ indicates the percentage of the flow on path p that is allowed to reach link e, taking capacity constraints into account. The (p, w)-element of $P \in \mathbb{R}^{\mathcal{P} \times W}$ indicates the probability of choosing path p for OD pair w. $d \in \mathbb{R}^{W}$ is the demand vector. By iteratively updating the link flow vector and the Ψ matrix, the flow vector will converge to the final solution.

Bundschuh e.a. (2006) propose a flow propagation method to determine congestion in a static assignment model. They call their model quasi-dynamic, since it does account for capacity limitations and spillback phenomena, but requires less computation time than a dynamic model. They assume a time period for which the static situation is valid. The method consists of two phases. In the first phase the traffic is propagated through the network, taking into account the link capacities. In the second phase the delay times are calculated.

The flow propagation is done in fractions. Dividing the flow in fractions reduces the influence of the order in which the routes and OD pairs are handled. The flow is propagated over the consecutive links of a route until the capacity of a certain link is reached. The extra flow on that link that cannot continue will be stored in the queue. Each link has a queue capacity, the maximal queue length in number of vehicles. If this amount is reached, and still more flow is propagated over this link, the surplus (i.e. spillback) is moved back to the preceding link(s) along the route.

In the second phase, the delay times are computed by determining the time it takes to empty the queues on the links in the network after the first phase while no extra flow is added. This is also done in steps by dividing the bottleneck link capacity into fractions.

In their model, Bundschuh e.a. introduce a permeability factor. This rule allows traffic to pass a queue that has spilt back, if the traffic follows a route that does not go through the link that caused the queue. This can happen for instance in the presence of separate turning lanes. The factor determines the fraction of traffic volume that is able to pass the queue.

It is not clear how Bundschuh e.a. determine the queue capacity. They assume that a vehicle always requires the same space within a congested link. However, it is known from the fundamental diagram (see section 3.2) that this is not true: the density (in vehicles/km) of a congested link is dependent of the flow (vehicles/hour). If the link is congested and there is only a little flow then the density is very high, but if there is only a little congestion and the flow is close to the link capacity, then the density is much lower. Surely Bundschuh e.a. take either the value of the critical density (density at capacity flow) or the jam density (fully congested, zero flow), or some value in between, but in general the queue length will not be accurate.

There is a tradeoff between the accuracy of the model and the computation time. Using more (smaller) fractions in the propagation phase and more fractions of capacity in the delay phase will increase the accuracy, but also increases the computation time. The algorithm of Bundschuh e.a. is implemented in the transport planning software package VISUM and is used in practice.

QBLOK (4Cast, 2009) is a traffic assignment algorithm that is used in practice in the Netherlands. Based on the capacity, travel demand and route choice behavior, the model assigns the traffic to the network. Route choice behavior is based on the first Wardrop principle. The model does account for congestion and spillback phenomena. The model does have some drawbacks. It can produce unrealistic travel times and strange route choices. The model uses a heuristic approach to reach an equilibrium situation by setting a predefined number of iterations and mixing the resulting link flows from different iterations according a certain distribution. It is not a mathematical model, but merely an algorithm that tries to find a satisfying result.

The Link Transmission Model (LTM) by Yperman (2007) is a dynamic network loading model designed for the dynamic traffic assignment problem. It describes the traffic propagation through a network in a realistic way (consistent with kinematic wave theory) and determines the link flows, densities and travel times. It does account for link capacities and spillback phenomena. The route flows from an existing route choice model are the input for the link transmission model. The model is built for the dynamic traffic assignment problem, but it is useful as a starting point for a static model.

2.4. Discussion

The attention for capacitated static traffic assignment that do account for queuing and spillback effects is scarce in the literature. In the table below, the different mathematical models that have been discussed in this chapter are categorized.

Mathematical Model	**Solution Method**	**Implementation**
Extended Beckmann User Equilibrium (with link capacity constraints)	Augmented Lagrangean dual with exterior penalty function	Larsson & Patriksson 1995 Nie e.a. 2004
	Interior penalty function	Nie e.a. 2004 Prashker & Toledo 2004
	Dynamic penalty function	Shahpar e.a. 2008
Stable Dynamics	n/a	Nesterov & De Palma 2000&2003
Flow propagation model	Propagate demand iteratively through the network	Bifulco & Crisalli 1998 Bundschuh e.a. 2006 Yperman 2007 (dynamic) 4Cast 2009

Table 1. Overview of different mathematical models that can deal with capacity constraints

Most papers about the side constrained or capacity constrained traffic assignment problem (Larsson & Patriksson 1995, Nie e.a. 2004, Prashker & Toledo 2004, Yang & Huang 2005, Shahpar e.a. 2008) follow the traditional approach and combine the route choice model and the determination of the link flows in the extended Beckmann optimization problem. The advantage of the (extended) Beckmann formulation is that it has a nice mathematical form, which can be solved using different solution techniques. However, those methods do not model the formation of queues and spillback (or blocking back) effects are not taken into account. It merely adds a time penalty for using a link that is filled to the capacity. However, since it is known from practice that congestion on a link can have consequences for the traffic flow on preceding links, these effects must be accounted for. Another disadvantage of the extended Beckmann formulation is that when a link does not have enough capacity, the travel time of that link will increase, instead of the travel time of the link(s) before that link. Besides that, most of these models apply travel time functions, which are in practice hard to determine and in general require time to calibrate to resemble acquired traffic counts.

There are some papers that try to obtain more realistic results. The Stable Dynamics model by Nesterov & De Palma uses only upper and lower bounds on the link flow and travel time respectively and utilizes the fundamental diagram to find a stable equilibrium. They propose adding an extra condition to deal with spillback. Unfortunately, in the rest of their work they do not use this condition; they just assume that spillback never occurs, so the effect of this condition in the model is unclear.
Bifulco & Crisalli (1998), Bundschuh e.a. (2006) and Yperman (2007) describe a flow propagation model (which contains no route choice model), taking a path flow vector coming from an existing route choice model as input. Bifulco & Crisalli describe a method to iteratively propagate the flow through the network by computing the fraction that is able to continue on the path given the link capacities. However, their model cannot handle spillback.
Bundschuh's model is a practical model that does account for spillback, but there are some drawbacks: they start building the queue in the bottleneck link instead of before the link. Besides that, they imply that the queue capacity of a link is constant. However, the queue capacity depends on the density of the queue.
QBlok does account for queuing and blocking back phenomena. However, this method is merely an algorithm using heuristics and if-then structures, and can result in strange route choices and unrealistic travel times. Since it is not a mathematical model, nothing can be said beforehand about the results either.
The dynamic Link Transmission Model (LTM) by Yperman (2007) does account for spillback effects, and describes the propagation of traffic flow through the network according to kinematic wave theory. Since the model is based on kinematic wave theory, it has a mathematical basis. Kinematic wave theory is considered as good method to model first order traffic flows.

In the figure on page 17, differences between some of the models are shown in a small hypothetical network. It gives an idea of the differences of the standard Beckmann model, the extended Beckmann model, the model Bifulco & Crisalli, the model of Bundschuh e.a. and QBlok. Furthermore it is shown how an ideal model should work.
Depicted is a simple network with seven links in series, with different capacities that are indicated by the width of the link. A traffic demand flow is sent through the network and it is shown what the link flows (widths of the colored parts) and travel times (color of the parts) will be in the different models.

If this is the only route on this network, the extended Beckmann model shall not have a feasible solution, because in some links there is not enough capacity. Therefore, it is assumed that there is an alternative route with unlimited capacity starting from the origin (upstream end of the first link) and ending at the destination (downstream end of the last link), with a very high travel time. In this way, this alternative route will only be chosen when there is no other option.

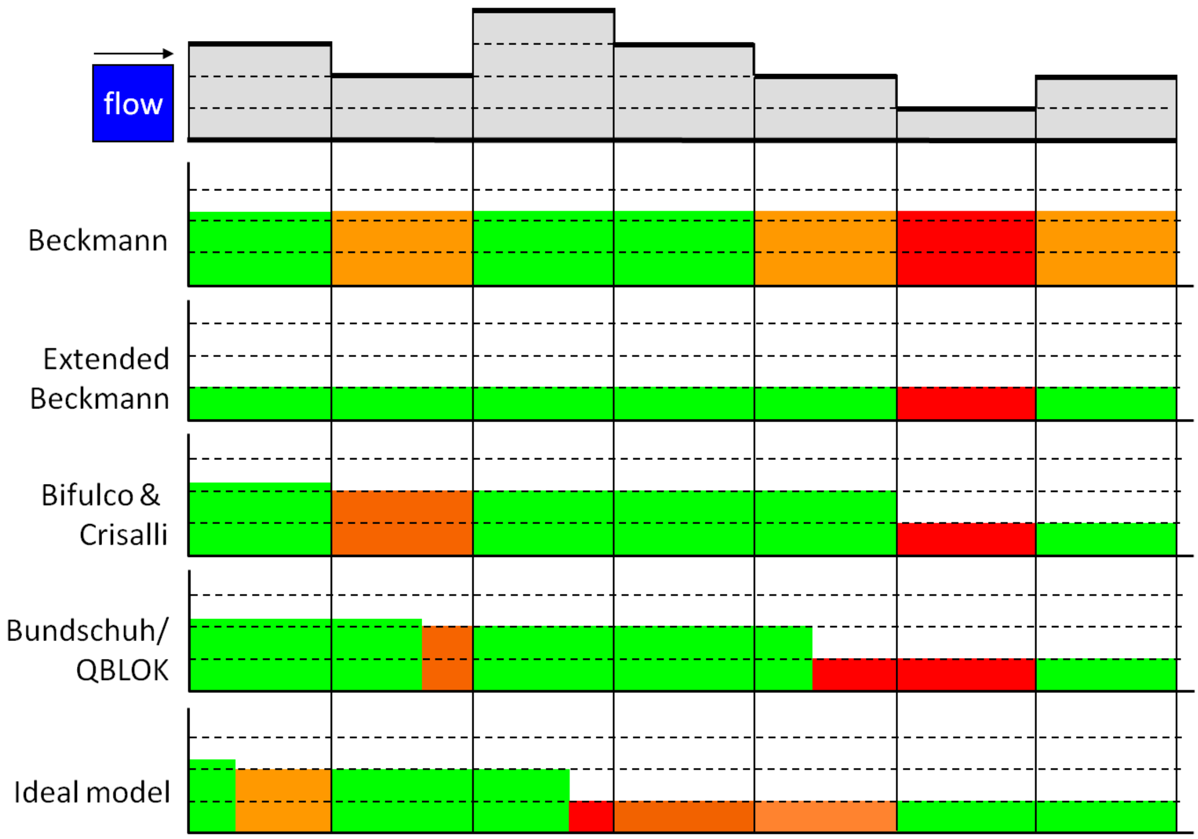


Figure 3. Comparison of traffic flows and speeds between different static models.

In the standard Beckmann model, where capacities are not taken into account, all link flows are equal to the input flow and are overestimated, and the travel times will be higher in the saturated links. In the extended Beckmann model, because of the low capacity of the sixth link, the flow on the whole route will be low and underestimated. The rest of the flow will be assigned to the alternative route. The travel time of the bottleneck link will be higher than in the other links. The third part of the figure shows the idea of the model Bifulco & Crisalli. The traffic flow pattern here is more realistic, but the queues are built inside the bottleneck links and spillback cannot occur. The fourth part shows the model by Bundschuh e.a. and the model QBLOK. Queues will appear inside the bottlenecks, and possibly spillback occurs. Both models work differently, but yield comparable results. It should be mentioned that unlike Bundschuh's model, QBLOK is merely a heuristic and not a complete model.

In the lowest part of the figure, a better way to model the congestion is shown. Queues are formed before the bottleneck and can possibly spill back to preceding links. The queue lengths are more accurate than in the model by Bundschuh e.a. Because the congestion modeling of this idea is similar to the method and the results of the dynamic Link Transmission Model, in the rest of this thesis a model that corresponds to this idea is formulated based on the LTM.

2.5. Conclusion

In this chapter different models and approaches from the literature were discussed and compared. The models are compared on the realism of the determination of the travel times and the mathematical or theoretical fundament the models are based on. The realism of the travel times is divided into four criteria (see Table 2).

Mathematical model	**Realistic determination of travel times**				**Theoretical fundament**
	Physical queuing	Queue location	Fundamental diagram	Spillback	
Extended Beckmann	-	-	-	-	+
Stable Dynamics	-	-	+	-	+
Flow propagation models:					
- Bifulco & Crisalli	-	-	-	-	+
- Bundschuh	+	-	-	+	+
- QBlok	+	-	-	+	-
- Link Transmission Model	+	+	+	+	+

Table 2. Summary of the comparison of the different approaches in literature on a number of criteria.

The Link Transmission Model is the best choice based on the criteria used. The LTM provides physical queuing and builds the queues before the bottleneck links and not inside them. It uses the complete fundamental diagram and spillback effects are accounted for. Since it is based on kinematic wave theory, it has a mathematical basis. However, since it is a dynamic model its higher computation time is a drawback. Therefore, it is chosen to derive the static variant of the Link Transmission Model in this thesis.

3. Link Transmission Model

The Link Transmission Model by Yperman (2007) is a dynamic network loading model that is used for the dynamic traffic assignment. The model realistically describes traffic propagation on a network. In this chapter the working of the Link Transmission Model (LTM) is described. In the first section the basic idea is explained. In the sections thereafter, the theory about the fundamental diagram, kinematic wave theory and shock wave theory is described shortly, to support further explanation of the model.

In the seventh section, an event-based, quasi-dynamic variant of the LTM is derived. It is called quasi-dynamic because it is assumed that the route flows are stationary, i.e. constant over time. Through this assumption, the model becomes less complex and far less calculations are needed. This adjusted, stationary Link Transmission Model is used as a step towards the Static Traffic Assignment with Queuing (STAQ) model, which will be the subject of chapter 4. In the last section an algorithm for the stationary LTM is presented.

3.1. Introduction to the LTM by Yperman

The original LTM is a dynamic network loading model. It is a simulation model that describes how the traffic propagates through the network over time in a realistic manner, consistent with the fundamental diagram and kinematic wave theory (see the next sections). It takes as input the time dependent route flows obtained by some existing route choice model. The LTM is the second component of a Dynamic Traffic Assignment model as illustrated in Figure 1 on page 1.

The LTM determines link volumes (i.e. number of vehicles) and travel times given the route flows. A route flow is loaded onto the network in the origin node and after some time leaves the network through its destination node.

The LTM counts the cumulative number of vehicles that have passed the beginning and the end of each link at time t. The number of vehicles that have passed location x at time t is $N(x,t)$. Location x^0 is the start of a link and location x^L is the end of the link, where L is the link length. So, the cumulative number of vehicles that have entered a link (cumulative inflow) at time t is $N(x^0,t)$. Similarly, $N(x^L,t)$ represents the cumulative number of vehicles that have left a link (cumulative outflow) at time t. By cumulative number of vehicles is meant the total number of vehicles since the start at time $t=0$. To simplify notation, $U(t)=N(x^0,t)$ is used for the cumulative inflow and $V(t)=N(x^L,t)$ for the cumulative outflow. It is assumed that there is no overtaking on a link (First In First Out behavior).

The cumulative in- and outflow can be visualized as in Figure 4. The difference between the time a vehicle entered and the time it left a link is the link travel time of that vehicle. The difference between in- and outflow is the number of vehicles on a link at a certain moment in time. So, if $U(t)=V(t)$, then there are no vehicles on this link. Note that $V(t) \leq U(t)$ always holds, since a vehicle has to enter a link before it can exit that link.

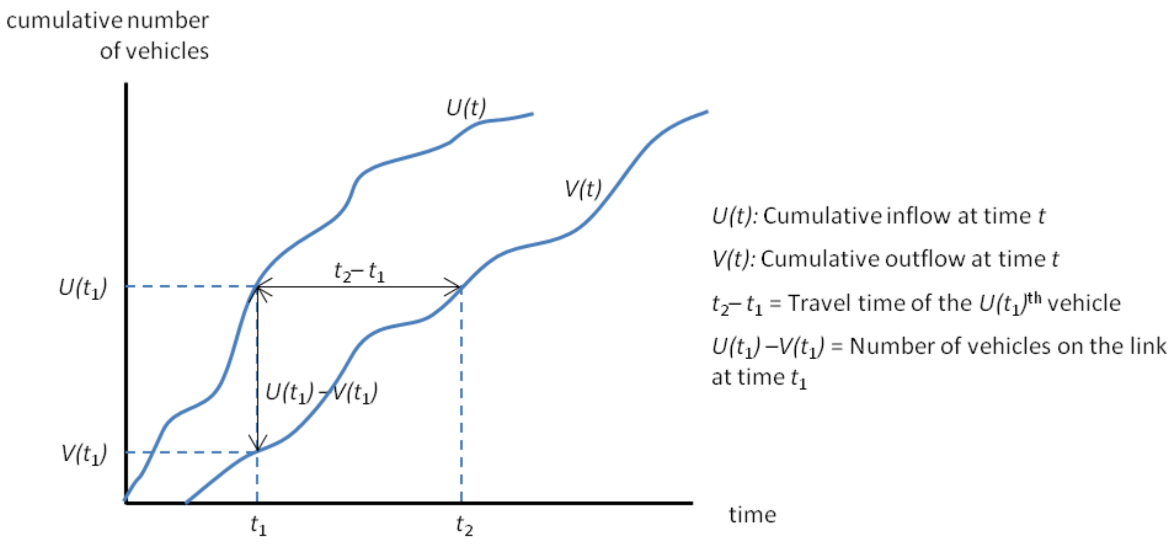


Figure 4. Example of cumulative in- and outflow curves of a link over time.

The LTM is a time based algorithm. For each time step (of fixed size), it determines for each link which amount of vehicles in that time step moves on to the next link, and adds this amount to the respective cumulative in- and outflows. To do this, the LTM uses the concept of sending, receiving and transition flows. The LTM can be divided into two parts: the link model and the node model.
In the node model it is determined which part of the traffic that wants to go towards a certain link is actually allowed to go to that link, based on the available capacities of the links connected to that node. In the link model, the traffic is propagated over a link in such a way that it is consistent with kinematic wave theory. A brief introduction to kinematic wave theory is given in section 3.3. Within kinematic wave theory, it is assumed that the fundamental relation between traffic flow (intensity), speed and density holds. This relation is described in a fundamental diagram. The fundamental diagram is the subject of the next section.

3.2. Fundamental diagram

A fundamental diagram describes the relation between the three important entities within traffic flow theory:

q traffic flow rate, intensity of the traffic (vehicles/hour)

k the density (German: konzentration) of the traffic (vehicles/km)

v the speed (velocity) of the traffic (km/hour)

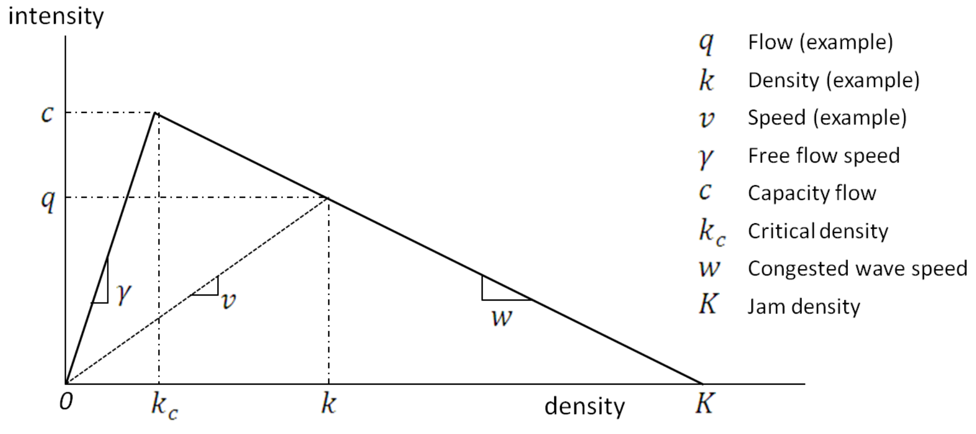
The relation $q = k \cdot v$ is known as the fundamental equation of traffic flow. This equation is based on the assumption that on average, drivers will drive the same under the same average conditions: at a certain speed v, they remain the same distance headway with respect to the next vehicle on the road. Each link can have a different fundamental diagram, but the diagram remains the same over time.
Three different diagrams can be visualized, all displaying the same information, but in a different form:

Speed – density: $v = v(k)$

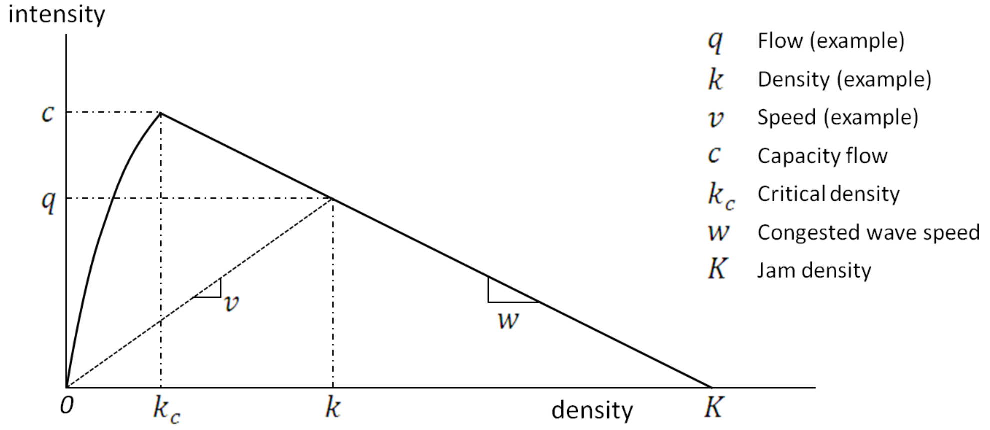
Speed – intensity: $v = v(q)$

Intensity – density: $q = q(k)$

The diagram showing the intensity – density relation is the most convenient for this thesis. In literature, different forms for the fundamental diagram have been proposed. One of the simplest forms is the diagram proposed by Newell (1993) and by Daganzo (1994) (Figure 5a). The intensity – density graph consists of two straight lines. From this graph, the speed v at a certain density k can be found, by determining the slope of the line from the origin to the point on the graph at that density (i.e. $v = q(k)/k$). The part of the diagram that lies to the left of the critical density k_c (increasing part of the diagram) is the uncongested part, in this part the traffic can flow with the free flow speed γ. At the critical density, the flow rate is equal to the capacity of the link, indicated by c. When the road traffic becomes more dense than k_c, both the flow that can go through the link and the speed of the flow decreases, until it reaches the jam density K. At that point, no flow is possible, so the speed is zero. Note that the diagram is determined by only three parameters: the free flow speed γ, the capacity c, and the jam density K. The slope w for the congested wave speed follows from these parameters by $w = c/(K - c/\gamma)$.



a



b

Figure 5. Two examples of a fundamental diagram: (a) Newells diagram and (b) Smulders' diagram.

Smulders (1990) introduced an similar form of this diagram, but he introduced a parabolic curve at the uncongested part (Figure 5b). This is a slightly more realistic way to model the speed in an uncongested link, because when the road becomes denser, speed will decrease: at the capacity flow, the speed is strictly less than the free flow speed ($= q'(0)$).

Like in Yperman (2007), Newells diagram (Figure 5a) is used in this thesis, because the diagram has some nice mathematical properties. Only two possible kinematic wave speeds exist (see section 3.5): the free flow speed γ for uncongested conditions (slope of first straigth line) and w for congested conditions (slope of second straight line).

3.3. Kinematic wave theory

The function $N(x,t)$ for the cumulative vehicle numbers at location x and time t is in general discontinuous, since the number of vehicles is discrete. Therefore it is assumed that a smooth approximation of the cumulative number of vehicles is available such that the function is twice differentiable. The approximation should coincide with the original function at the positions where a vehicle enters or exits the link, i.e. the 'steps' in the function. In this way the partial derivatives of $N(x,t)$ can be determined. The partial derivative with respect to t can be interpreted as the instantaneous flow (in vehicles per time unit) at point (x,t):

$$\frac{\partial N(x,t)}{\partial t} = q(x,t)$$

The partial derivative with respect to x can be interpreted as the density (in vehicles per kilometer) at point (x,t):

$$\frac{\partial N(x,t)}{\partial x} = -k(x,t)$$

The density has a negative sign because for increasing x, $N(x,t)$ is decreasing. It is assumed that our smooth approximation of $N(x,t)$ has continuous second partial derivatives, so the identity

$$\frac{\partial^2 N(x,t)}{\partial x \partial t} = \frac{\partial^2 N(x,t)}{\partial t \partial x}$$

yields

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 \qquad (6)$$

This equation is known as the conservation law. It ensures that changes in the rates of the flow and density over time and space cannot cause vehicles to appear or disappear.
In kinematic wave theory, it is assumed that traffic behaves as described by a fundamental diagram and that the flow can be taken as a function of the density. This means that $q(x,t)$ can be replaced with $Q\big(k(x,t)\big)$ in equation (6):

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial Q\big(k(x,t)\big)}{\partial x} = \frac{\partial k(x,t)}{\partial t} + \frac{dQ\big(k(x,t)\big)}{dk}\frac{\partial k(x,t)}{\partial x} = 0 \tag{7}$$

Now the conservation law is formulated with only one independent variable, the density k. In the following, Q' is used to represent the value of the derivative of Q, $dQ\big(k(x,t)\big)/dk$. Since a fundamental diagram with two linear segments is used (Figure 5a), only two different values for Q' can occur: the free flow wave speed γ and the congested wave speed w. As long as Q' is constant, solutions of the partial differential equation (7) are of the form:

$$k(x,t) = F(x - Q' \cdot t)$$

Indeed,

$$\frac{\partial k(x,t)}{\partial t} = F'(x - Q' \cdot t) \cdot (-Q') \text{ and } \frac{\partial k(x,t)}{\partial x} = F'(x - Q' \cdot t) \cdot 1$$

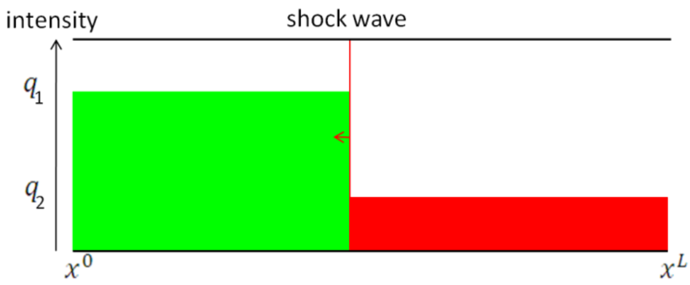
imply

$$\frac{\partial k(x,t)}{\partial t} + Q' \cdot \frac{\partial k(x,t)}{\partial x} = F'(x - Q' \cdot t) \cdot (-Q') + Q' \cdot F'(x - Q' \cdot t) = 0$$

This means that in the $t - x$ plane the density is constant on straight lines with slope Q'. These lines are called characteristics or waves. All along such a characteristic line, the traffic state conditions are the same. A traffic state can be seen as a certain combination of the traffic flow and density. So, a traffic state moves along such a wave through the $t - x$ plane. At one position in the $t - x$ plane (a certain combination of place and time) there can be only one traffic state. When two different traffic states intersect with each other, a shock wave appears.

3.4. Shock waves

When there is a change in traffic conditions, such as a change in the inflow or outflow caused by a decrease in capacity, an imaginary boundary is established in the $t - x$ plane. This boundary indicates a change from one traffic state to another. Such boundaries are called shock waves. Shock waves can move with or against the direction of the traffic, depending on the change in traffic states. This is explained with an example. In Figure 6 a link with a shock wave is shown, next to its associated fundamental diagram. There is a change from traffic state 1 (density k_1, flow q_1) to traffic state 2 (density k_2, flow q_2), which causes a queue to grow in the link. The shock wave indicates the start of the queue. When the shock wave reaches the upstream end of the link (x^0), the queue shall spill back to the preceding links.



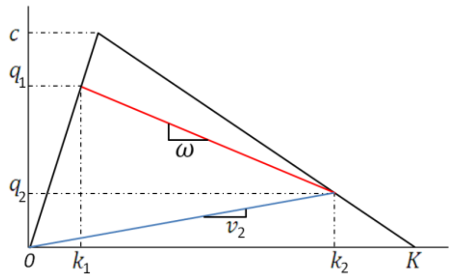


Figure 6. Example of a backward shock wave, with its associated fundamental diagram.

The height of the colored parts indicate the intensity of the flow. The green part indicates free flow conditions (traffic state 1) and the red part indicates congested conditions (traffic state 2), the speed of the traffic will be lower there. Suppose the change of traffic states is initiated at time t_1 at location x^L. The shock wave reaches location x^0 a certain time later, say at time t_2. The difference in number of vehicles on the link between the moment that the shock wave is initiated and the moment it reaches the other link end can be found by:

$$\Delta L \cdot \Delta k = (x^0 - x^L)(k_1 - k_2)$$

Alternatively, this number can be found through the change in flow rates:

$$\Delta t \cdot \Delta q = (t_2 - t_1)(q_1 - q_2)$$

Since $\Delta L \cdot \Delta k = \Delta t \cdot \Delta q$ and the shock wave travels the distance ΔL in time Δt, the speed ω of the shock wave is found as follows:

$$\omega = \frac{\Delta L}{\Delta t} = \frac{q_2 - q_1}{k_2 - k_1} \tag{8}$$

This speed is also found by determining the slope of the line between the two points on the fundamental diagram (see Figure 6). In this example ω is negative, because the shock wave travels against the direction of the traffic.

In Figure 7, a $t - x$ plane is displayed of a link a. x_a^0 is the upstream link end, $x_a^{L_a}$ is the downstream link end, where L_a is the length of link a. At time t_1 the outflow rate (at $x_a^{L_a}$) decreases from q_1 to q_2, caused by a capacity reduction. The lines with slope γ_a are traffic waves within free flow conditions. Along these waves, the intensity or flow rate is q_1 and the density is k_1. For such a wave, it takes L_a/γ_a to reach the end of the link if it does not intersect with another state. The lines with slope w_a are traffic waves within congested conditions. Along these waves the intensity is q_2 and the density is k_2, and it takes $-L_a/w_a$ to reach the beginning of the link if it does not intersect with another state. The red line with slope ω indicates the intersection of the two different traffic states, and shows the movement of the shock wave through time and space. It takes $-L_a/\omega = t_2 - t_1$ time for the shock wave to reach the beginning of the link. The diagram also shows some vehicle trajectories. In the free flow part, these trajectories coincide with the traffic waves, because the vehicles travel with wave speed γ_a. If a vehicle reaches the shock wave, it transfers to the congested area and continues to travel with speed v_2 (see also the fundamental diagram of Figure 6).

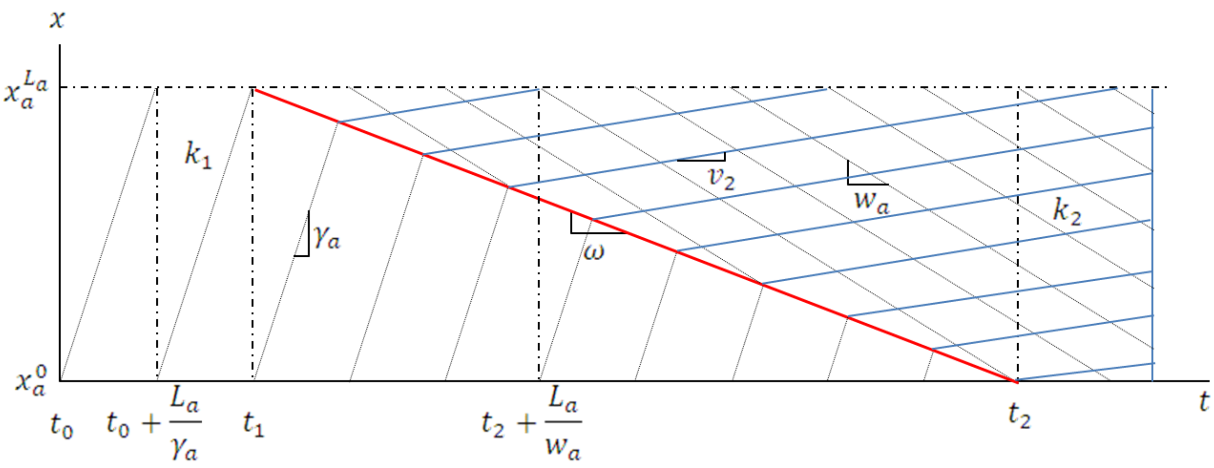


Figure 7. Example of a $t - x$ plane with a shock wave as in Figure 6.

Note that it is assumed that a transition from one traffic state to another does not take time, which implies that vehicles do not accelerate or decelerate. Vehicles make a jump in their driving speed. This assumption simplifies the model greatly, and since the purpose is to model the first order traffic flow and not single vehicles, it is accepted.

There can be many shock waves on the links in a network. Each time a shock wave reaches a link end, it results in new shock waves in the other links connected to this node. By keeping track of all the shock waves in the network, the situation of the traffic flows and densities on the links in the network after a certain time period, for instance one hour, can be calculated.

The Link Transmission Model does model the traffic flows according to the shock waves in the network. However, by using Newell's simplified theory of kinematic waves (see next section), the LTM only needs to keep track of the cumulative number of vehicles that have entered and left each link, and explicitly keeping track of all the shock waves is no longer needed.

3.5. Newell's simplified theory of kinematic waves

As explained before, when using a triangular shaped fundamental diagram, a traffic state can travel through a link with just two possible wave speeds. A traffic state within free flow conditions travels from the upstream boundary (beginning) of link a to the downstream boundary (end) of the link with free flow speed γ_a. A congested traffic state travels from the downstream boundary of link a to the upstream boundary with speed w_a (w_a is negative, so it travels against the direction of traffic). See also the $t - x$ plane of Figure 7.

A free flow traffic state traveling from the upstream link end following a free flow traffic wave reaches the end of the link L_a/γ_a time units later, if it does not intersect with another state. Because vehicle trajectories coincide with traffic waves within free flow conditions, the change in cumulative vehicle numbers is zero:

$$N(x_a^0, t) - N\left(x_a^{L_a}, t + \frac{L_a}{\gamma_a}\right) = U_a(t) - V_a\left(t + \frac{L_a}{\gamma_a}\right) = 0$$

$U_a(t)$ is defined as the cumulative number of vehicles that have entered link a at time t (cumulative inflow), and $V_a(t)$ is defined as the cumulative number of vehicles that have left link a at time t

(cumulative outflow). So, when the link is in free flow conditions (when the density is smaller than the critical density, i.e. no congestion), the cumulative number of vehicles that have left the link is equal to the cumulative number of vehicles that have entered the link L_a/γ_a time units earlier:

$$V_a(t) = U_a\left(t - \frac{L_a}{\gamma_a}\right) \tag{9}$$

In Figure 10 the graph of the cumulative vehicle numbers within free flow conditions is shown, with inflow rate q_1.

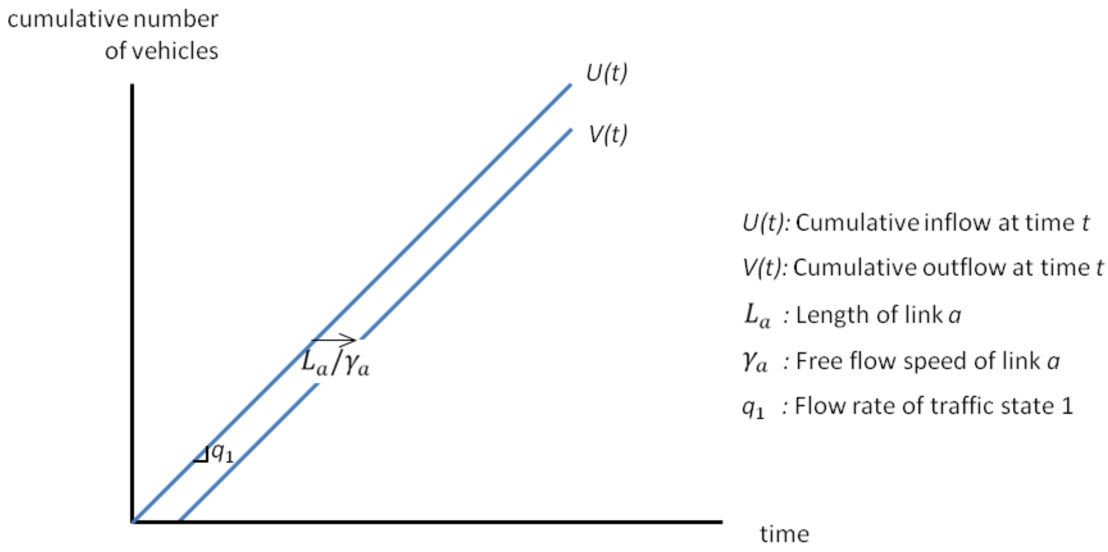


Figure 8. Cumulative vehicle numbers within free flow conditions.

A congested traffic state traveling from the downstream link end following a congested traffic wave reaches the beginning of the link $-L_a/w_a$ time units later, if it does not intersect with another state. With Green's Theorem (1828) it can be shown that the change in cumulative vehicle numbers is K_aL_a, where K_a is the jam density of link a (see also Yperman (2007)). Here C is a straight line in the $t-x$ plane between the points (x_a^0, t) and $\left(x_a^{L_a}, t + \frac{L_a}{w_a}\right)$ along which q and k are constant; both points represent the same congested traffic state (q, k).

$$N(x_a^0, t) - N\left(x_a^{L_a}, t + \frac{L_a}{w_a}\right) = \int_C dN(x,t)$$

$$= \int_C \frac{\partial N(x,t)}{\partial t} dt + \frac{\partial N(x,t)}{\partial x} dx = \int_C qdt - kdx$$

$$= \left(qt - q\left(t + \frac{L_a}{w_a}\right)\right) - (k \cdot 0 - k \cdot L_a)$$

$$= -q \cdot \frac{L_a}{w_a} + k \cdot L_a = L_a \cdot \left(-\frac{q}{w_a} + k\right)$$

$$= K_aL_a$$

The last step is found using the fundamental diagram (see Figure 5a): $(K_a - k) \cdot -w_a = q$ holds for a congested traffic state (q, k), which yields $K_a = k - q/w_a$. So, within congested conditions, the cumulative number of vehicles that entered the link at time t is equal to the cumulative number of vehicles that have left the link $-L_a/w_a$ time units earlier plus $K_a L_a$:

$$U_a(t) = V_a\left(t + \frac{L_a}{w_a}\right) + K_a L_a \tag{10}$$

Suppose that as in Figure 7, there is a traffic flow rate of q_1 on a link. At time t_1 the outflow capacity of the link is reduced to q_2. This will result in congestion on the link, because the current traffic flow rate in the link is $q_1 > q_2$. Therefore, from time t_1 the graph of the cumulative outflow will have slope q_2, and from time $t_1 - L_a/w_a$ the cumulative inflow will also have slope q_2, with $K_a L_a$ more vehicles (see Figure 9).

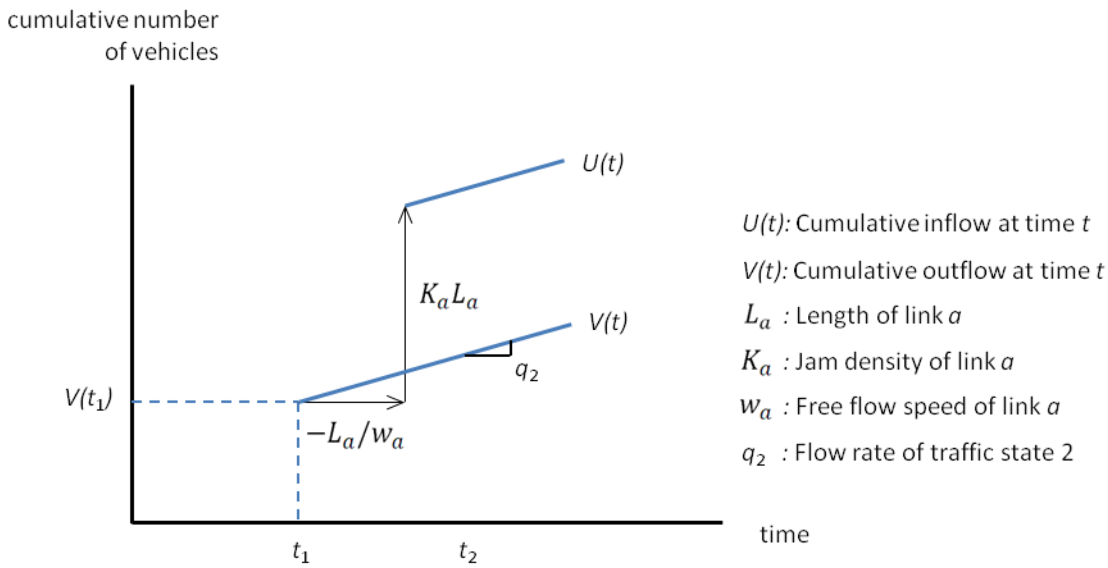


Figure 9. Cumulative vehicle numbers within congested conditions.

This results in a double valued solution for the cumulative inflow and outflow, by Figure 8 and Figure 9. The unique solution is achieved by taking the lower envelop of this double valued solution (see Figure 10). Note that in the unique solution the inflow rate changes at time t_2 from q_1 to q_2, just as found in Figure 7. This solution is found without explicitly calculating the shock wave. This approach works even with multiple shock waves on a link.

The following lemma shows that the approach above yields the same outcome as calculating the shock wave arrival time explicitly as was done in section 3.4. It shows that the time t_2 found in Figure 7 is the same as the t_2 in Figure 10.

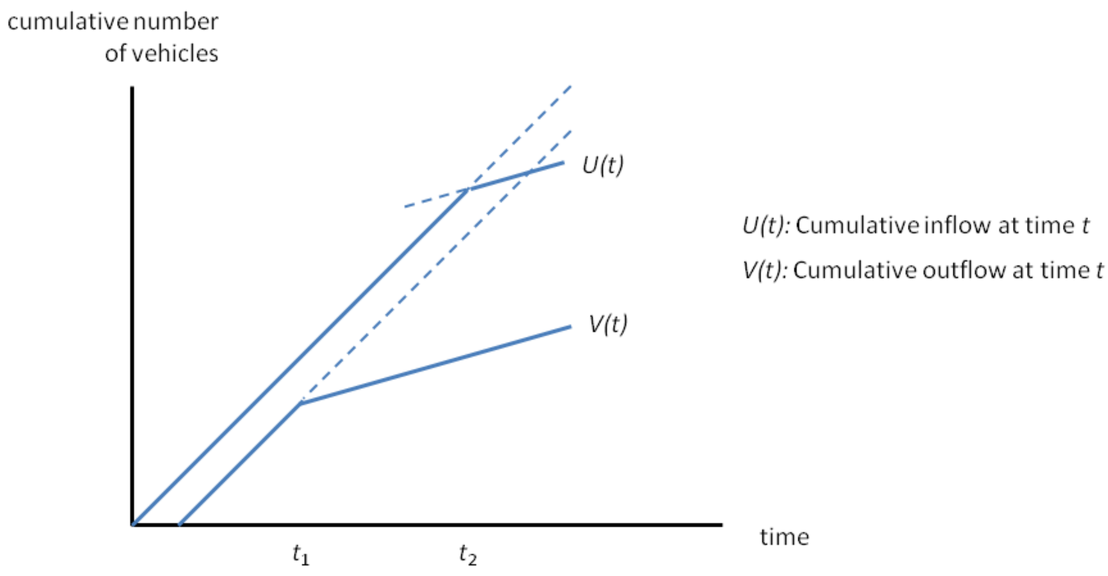


Figure 10. The lower envelop of the double valued solution of Figure 8 and Figure 9.

Lemma

Consider a link with traffic state (q_1, k_1) on the whole link. At time t_1 the outflow rate changes to $q_2 < q_1$, with corresponding (congested) density k_2. Calculating the arrival time of the shock wave explicitly through the shock wave speed yields the same result as calculating it via Newell's simplified theory of kinematic waves.

Proof

1. Explicit shock wave arrival time. Following equation (8), the shock wave travels with speed $\omega = \frac{q_2 - q_1}{k_2 - k_1}$, so it reaches the upstream link end $-L_a/\omega$ time units after t_1 (ω is negative):

$$t_2 = t_1 - \frac{L_a}{\omega}$$

2. Newell's simplified theory of kinematic waves. The arrival time t_2 of the shock wave can be calculated following the approach derived from Newell's simplified theory of kinematic waves. The change in the inflow rate is caused by the arrival of the shock wave at the upstream link end. So, t_2 equals the time at which the graph of the cumulative inflow from Figure 10 changes direction, i.e. the time that the graph of the cumulative inflow from Figure 8 crosses the graph of the cumulative inflow from Figure 9:

$$U_{\text{Figure 8}}(t_2) = U_{\text{Figure 9}}(t_2)$$

The cumulative inflow at time t_2 of Figure 8 is simply $t_2 \cdot q_1$. The cumulative inflow at time t_2 of Figure 9 is determined in three parts. Until t_1 the cumulative outflow is $(t_1 - L_a/\gamma_a) \cdot q_1$, then a step of $K_a L_a$ is taken towards the cumulative inflow. Then there is another $\left(t_2 - (t_1 - L_a/w_a)\right)$ time of inflow rate q_2. This yields:

$$t_2 \cdot q_1 = \left(t_1 - \frac{L_a}{\gamma_a}\right) q_1 + K_a L_a + \left(t_2 - \left(t_1 - \frac{L_a}{w_a}\right)\right) q_2$$

Since $\gamma_a = v_1$ and $k_1 = q_1/v_1$ it follows that:

$$t_2 \cdot q_1 = t_1 \cdot q_1 - L_a k_1 + K_a L_a + t_2 \cdot q_2 - t_1 \cdot q_2 + \frac{L_a}{w_a} q_2$$

Furthermore, since from the fundamental diagram it is known that $(K_a - k_2) \cdot -w_a = q_2$, it follows that $\frac{L_a}{w_a} \cdot q_2 = L_a(k_2 - K_a)$, so:

$$-t_1 \cdot q_1 - t_2 \cdot q_2 + t_1 \cdot q_2 + t_2 \cdot q_1 = -L_a k_1 + K_a L_a + L_a k_2 - K_a L_a$$

$$(t_1 - t_2)(q_2 - q_1) = L_a(k_2 - k_1)$$

$$t_2 = t_1 - L_a \cdot \frac{k_2 - k_1}{q_2 - q_1}$$

Since $\omega = (q_2 - q_1)/(k_2 - k_1)$, this is the same result as above. ■

Kinematic wave theory is described in more detail by e.g. Daganzo (1997) and Newell (1993).

3.6. Formulation of the Link Transmission Model

Yperman uses equations (9) and (10) in his formulation of sending and receiving flows. In the following, the Link Transmission Model (LTM) by Yperman (2007) is described.

Input for the LTM:

- Network (V, E) including for each link $a \in E$:
  - $c_a > 0$ link capacity (vehicles/hour)
  - $L_a > 0$ link length (km)
  - $K_a > 0$ jam density, density at which flow is no longer possible (vehicles/km)
  - $\gamma_a > 0$ forward (free flow) wave speed (km/hour)
  - $w_a < 0$ backward (congested) wave speed (km/hour)

- The fundamental diagram of link a, determined by the parameters c_a, K_a and γ_a (see Figure 5a). w_a follows from these parameters by $w_a = c_a/(K_a - c_a/\gamma_a)$.

- $\mathcal{P}$: Set of used routes (or paths)

- $f_p(t) > 0$: Path flow rates for each path $p \in \mathcal{P}$ at time t, the traffic flow per hour assigned to that path (vehicles/hour). Since LTM is a dynamic model, path flows can change over time. $f_p(t)$ is the rate at which vehicles enter the network at the upstream link end of the first link of path p at time t. Path flows are assumed to be given, achieved by some route choice model.

- Δt, the fixed time step in the algorithm. Δt should be smaller than the smallest link travel time. This requirement is known as the Courant-Friedrichs-Lewy (CFL) condition:

$$\Delta t \leq \frac{L_a}{\gamma_a} \qquad \forall a \in E$$

To indicate that two links are connected to each other, or more precisely that the downstream end of link a is connected to the same node as the upstream end of link b, the notation $a \rightarrow b$ is used.

Variables for the LTM:

$U_a(t)$	The cumulative number of vehicles that have entered link a at time t.
$V_a(t)$	The cumulative number of vehicles that have left link a at time t.
$S_a(t)$	The sending flow of link a at time t (maximum number of vehicles that could leave the downstream end of this link during $[t, t + \Delta t]$, if this link end were connected to a traffic reservoir with an infinite capacity).
$S_{ab}(t)$	The directional sending flow from link a to link b at time t, the part of $S_a(t)$ that wants to go to link b. It is assumed that $a \rightarrow b$.
$R_a(t)$	The receiving flow of link a at time t (maximum number of vehicles that could enter the upstream end of this link during $[t, t + \Delta t]$, if a traffic reservoir with an infinite traffic demand were connected to this link end).
$G_{ab}(t)$	The transition flow of link a to link b at time t (number of vehicles that are actually transferred from link a to link b during $[t, t + \Delta t]$). It is assumed that $a \rightarrow b$.

The LTM basically consists of two parts, the link model and the node model. In the link model the traffic is propagated through the links of the network by determining the sending and receiving flows of each link. The sending and receiving flows indicate the amount of vehicles that are willing to exit and able to enter a certain link, respectively. In the node model, the transition flows are determined.

The sending flow of link a at time t is the maximum amount of vehicles that could leave the downstream end of this link during $[t, t + \Delta t]$, if there are no capacity constraints downstream this link. This means there are free flow conditions in the link. The upper bound for the sending flow is the difference in cumulative outflows at time t and $t + \Delta t$, where $V_a(t + \Delta t)$ can be obtained using equation (9) on page 26:

$$S_a(t) \leq V_a(t + \Delta t) - V_a(t) = U_a\left(t + \Delta t - \frac{L_a}{\gamma_a}\right) - V_a(t) \qquad (11)$$

The sending flow is also constrained by the capacity of the link:

$$S_a(t) \leq c_a \cdot \Delta t \qquad (12)$$

So, the sending flow is the maximum flow taking into account equations (11) and (12):

$$S_a(t) = \min\left(U_a\left(t + \Delta t - \frac{L_a}{\gamma_a}\right) - V_a(t); c_a \cdot \Delta t\right) \tag{13}$$

The receiving flow of link a at time t is the maximum amount of vehicles that could enter the upstream end of this link during $[t, t + \Delta t]$, if there is infinite traffic willing to enter this link. This means that the inflow is only constrained by the downstream link end, in case of congestion. The upper bound for the receiving flow is the difference in cumulative inflows at time t and $t + \Delta t$, where $U_a(t + \Delta t)$ can be obtained using equation (10) on page 27:

$$R_a(t) \leq U_a(t + \Delta t) - U_a(t) = V_a\left(t + \Delta t + \frac{L_a}{w_a}\right) - U_a(t) + K_a L_a \tag{14}$$

The receiving flow is also constrained by the capacity of the link:

$$R_a(t) \leq c_a \cdot \Delta t \tag{15}$$

So, the receiving flow is the maximum flow taking into account equations (14) and (15):

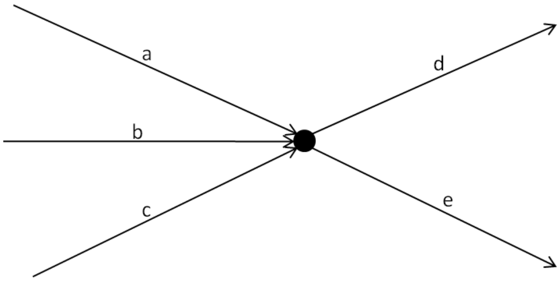
$$R_a(t) = \min\left(V_a\left(t + \Delta t + \frac{L_a}{w_a}\right) - U_a(t) + K_a L_a; c_a \cdot \Delta t\right) \tag{16}$$

In the node model, the transition flows are determined. The transition flows indicate which part of the sending and receiving flows can actually be sent and received, given the situation at that node.

$$G_{ab}(t) = \min_{\substack{b' \in E, \\ a \to b'}}\left(\frac{R_{b'}(t)}{\sum_{\substack{a' \in E, \\ a' \to b}} S_{a'b'}(t)} \cdot S_{ab}(t);\ S_{ab}(t)\right) \tag{17}$$

In this node model, the outflow of all incoming links of a node will be reduced if there is a capacity problem at that node. The total flow that want to flow to a outgoing link b is the sum of the directional sending flows $S_{ab}(t)$. However, at most there can flow as much as the receiving flow $R_b(t)$ to this link. The link with the smallest ratio between the receiving flow and the sum of the directional sending flows determines the factor of the sending flows is allowed to go to the next link at this node.

So, it is assumed that at a junction with a capacity problem, the outflow of all incoming links is reduced. Consider an intersection as Figure 11, a node with three incoming links (a, b, c) and two outgoing links (d, e). The directional sending flows for a certain time t are shown in Table 3. All links have a capacity of 2000 vehicles per hour, and at this time t all receiving flows are equal to the capacity.



S_{ij}	d	e
a	900	0
b	600	1200
c	0	1200

Figure 11. Example of a node with a capacity problem.

Table 3. Directional sending flows for the junction of Figure 11.

Because there is a flow of 2400 vehicles per hour that wants to go to link e, there is a capacity problem. Following the node model, the transition flows are determined by taking the directional sending flows and multiplying this with the smallest ratio at this node, which is 2000/2400. So, the transition flows are as follows: $G_{ad} = 750$, $G_{bd} = 500$, $G_{be} = G_{ce} = 1000$. Note that even the outflow from link a to link d is decreased, while there is no flow from a to e. However, in this node model it is assumed that when an intersection is congested, all traffic streams are influenced with the same factor.

This node model is chosen for its simplicity, as the formulation is very straightforward. In the paper by Tampère e.a. (2011) this node model is criticized. They show with a numerical example that the total flow over the node with respect to the sending flows is not always maximized. Furthermore, the invariance principle may be violated, which means that a situation can arise in which a queue alternatively grows and dissolves while the boundary conditions (i.e. inflows of the incoming links, outflows of the outgoing links) are constant. The impact of this flaw is not clear and further research on this subject is needed.
A different node model formulation could be implemented in the link transmission model as long as it is applicable to a general intersection, it uses sending- and receiving flows, and the conservation of vehicles and the conversation of turn fractions is guaranteed. Obviously, the flows also need to be non-negative and the demand and supply constraints must be satisfied. The conservation of turn fractions means that the fraction of the sending flow to a certain link is equal to the fraction of the actual flow to that link:

$$tf_{ab} = \frac{S_{ab}}{S_a} = G_{ab} / \sum_{\substack{b' \in E, \\ a \to b'}} G_{ab'} \qquad \forall a, b \in E, \; a \to b \tag{18}$$

If there is FIFO (first in first out) behavior on the link this requirement is automatically satisfied.

The following sets are used in the algorithm:
I_n is the set of incoming links into node n, i.e. all links that have node n as the downstream link end.
J_n is the set of outgoing links from node n, i.e. all links that have node n as the upstream link end.
For example, for the node in Figure 11, this means that $I_n = \{a, b, c\}$ and $J_n = \{d, e\}$.

The algorithm of the original LTM is as follows. For details, see Yperman (2007).

For each time interval Δt:
For each node n:

1. For each $a \in I_n$ determine the sending flow $S_a(t)$ using (13), and for each $b \in J_n$ determine the receiving flow $R_b(t)$ using (16).

2. Determine the transition flows $G_{ab}(t)$ from link a to link b using (17) for all $a \in I_n$, $b \in J_n$.

3. Update cumulative vehicle numbers:

$$V_a(t + \Delta t) = V_a(t) + \sum_{b \in J_n} G_{ab}(t) \qquad \forall a \in I_n$$

$$U_b(t + \Delta t) = U_b(t) + \sum_{a \in I_n} G_{ab}(t) \qquad \forall b \in J_n$$

3.7. The stationary Link Transmission Model using flow rates

In the rest of this thesis, it is assumed that the traffic demand is stationary, as in a static traffic assignment model. In this way it is possible to derive the static variant of the LTM. This means that during the whole studied time period the Origin-Destination (OD) demand matrix remains the same, so there is a constant inflow rate onto the first link of each path. This is a static approximation or average of the dynamic demand during a certain time period. In Figure 12 an example of the dynamic demand on a link during the morning peak hours is shown. The stationary approximation will be somewhere between the top and the bottom of the peak. The best way to choose the approximation is not within the scope of this thesis; it is assumed that the stationary demand is given.

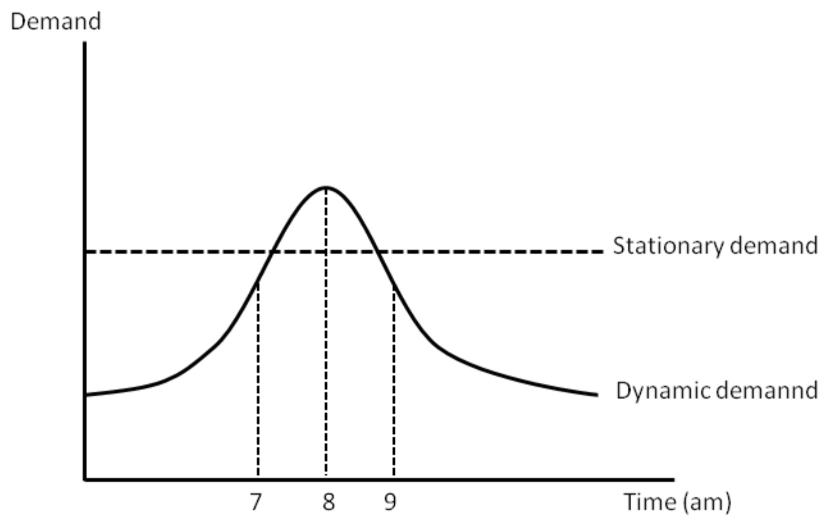


Figure 12. Example of the dynamic demand during morning peak hours, and a stationary approximation.

Because the OD demand is constant, also the assigned path flows achieved from the route choice model are constant. So, the inflow rate (or intensity) of each first link of all routes is constant. However, because of congestion and spillback effects, this is not true for all links in general. In the following, the sending and receiving flows will be defined in vehicles per hour, at a certain time, instead of a number of vehicles in a certain time interval. Thus, they become sending and receiving flow rates, and also in- and outflow rates. The sending- and receiving flow rates of a node can only change when an (implicit) shock wave reaches this node, for instance when a queue has filled the whole link and is about to spill back. These moments when there is a change in traffic states at a node are called events.

The formulas from the original Link Transmission Model are derived for the stationary case. The event times can be calculated, without iterating for each time step. The time that the next sending or receiving flow rate changes can be determined immediately, instead of calculating the sending and receiving flows for each time step Δt. In Figure 13 an example of the cumulative in- and outflows is given. The slope of the cumulative in- and outflows correspond to the in- and outflow rate respectively. The changes in the slope correspond to the event times. It is assumed that these changes do not take time.

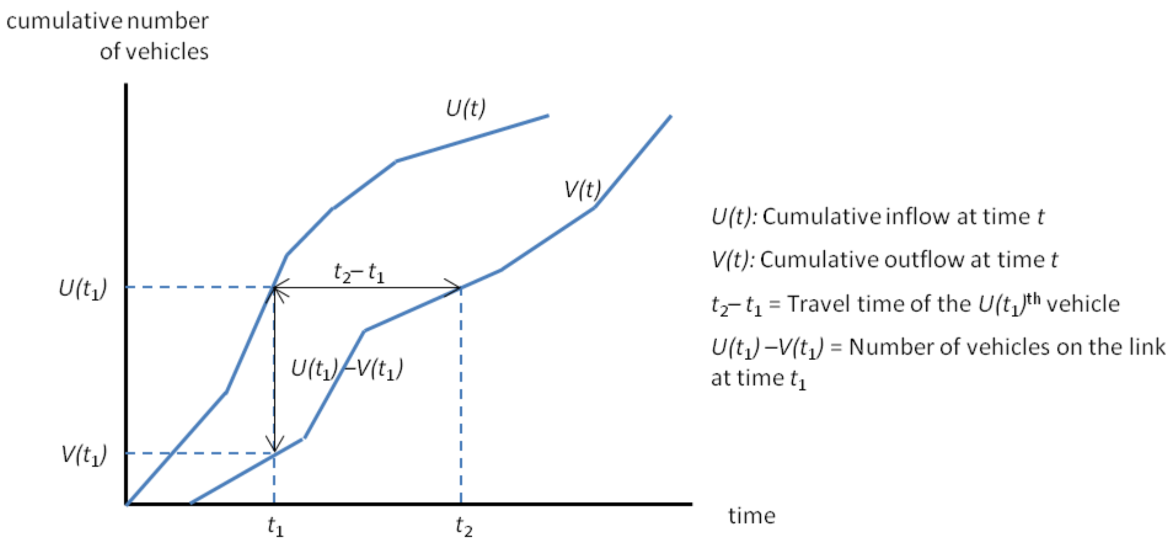


Figure 13. Example of cumulative in- and outflow curves of a link over time.

Again the link travel time of a vehicle can be derived by the difference between the time that the vehicle entered and the time that it left the link. The difference between in- and outflow is the number of vehicles on a link at a certain moment in time.

Largely the same notation and set of variables as in the original LTM is used. The differences are that Δt is no longer used, that f_p is not time dependent, and that the sending, receiving and transition flows are in vehicles per hour instead of number of vehicles per Δt. The model is performed for a predefined period, for instance one hour, and from the cumulative in- and outflows at the end of that period the travel times can be derived.

Input for the stationary LTM:

- Network (V, E) including for each link $a \in E$:
  - $c_a > 0$ link capacity (vehicles/hour)
  - $L_a > 0$ link length (km)
  - $K_a > 0$ jam density, density at which flow is no longer possible (vehicles/km)
  - $\gamma_a > 0$ forward (free flow) wave speed (km/hour)
  - $w_a < 0$ backward (congested) wave speed (km/hour)

- The fundamental diagram of link a, determined by the parameters c_a, K_a and γ_a (see Figure 5a). w_a follows from these parameters by $w_a = c_a/(K_a - c_a/\gamma_a)$.

- $\mathcal{P}$: Set of used routes (or paths)

- $f_p > 0$: Path flow rates for each path $p \in \mathcal{P}$, the traffic flow per hour assigned to that path (vehicles/hour). f_p is the rate at which vehicles enter the network at the upstream link end of the first link of path p. Path flow rates are assumed to be given, achieved by some route choice model.

Variables for the stationary LTM:

$U_a(t)$	The cumulative number of vehicles that have entered link a at time t.
$V_a(t)$	The cumulative number of vehicles that have left link a at time t.
$u_a(t)$	The inflow rate of link a at time t (vehicles/hour).
$v_a(t)$	The outflow rate of link a at time t (vehicles/hour).
$s_a(t)$	The sending flow rate of link a at time t (maximum outflow rate at time t, if the downstream end of this link were connected to a traffic reservoir with an infinite capacity; in vehicles/hour).
$s_{ab}(t)$	The directional sending flow rate from link a to link b at time t, the part of $s_a(t)$ that wants to go to link b (vehicles/hour). It is assumed that $a \rightarrow b$.
$r_a(t)$	The receiving flow rate of link a at time t (maximum inflow rate at time t, if a traffic reservoir with an infinite traffic demand were connected to the upstream end of this link; in vehicles/hour).
$g_{ab}(t)$	The transition flow rate of link a to link b at time t (highest actual possible flow rate between link a to link b at time t; in vehicles/hour). It is assumed that $a \rightarrow b$.

In the following, the fundamental rule of calculus is used to take the derivatives of the cumulative number of vehicles $U_a(t)$ and $V_a(t)$ to achieve the respective in- and outflow rates (in vehicles per hour) at time t. Using equation (9) on page 26, the relation of the in- and outflow rates in free flow conditions is obtained:

$$v_a(t) = V_a{}'(t) = U_a{}'\left(t - \frac{L_a}{\gamma_a}\right) = u_a\left(t - \frac{L_a}{\gamma_a}\right) \tag{19}$$

Similarly, using equation (10) on page 27 yields the relation within congested conditions:

$$u_a(t) = U_a{}'(t) = V_a{}'\left(t + \frac{L_a}{w_a}\right) = v_a\left(t + \frac{L_a}{w_a}\right) \tag{20}$$

The sending flow $S_a(t)$ in the original LTM is the maximum number of vehicles that can leave the link in a time period $[t, t + \Delta t]$, where Δt is fixed. Dividing this number by Δt yields the average sending flow rate in this time period. Then letting $\Delta t \to 0$, the sending flow rate at time t is obtained.

$$s_a(t) = \lim_{\Delta t \to 0} \frac{S_a(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\min\left(U_a\left(t + \Delta t - \frac{L_a}{\gamma_a}\right) - V_a(t); c_a \cdot \Delta t\right)}{\Delta t}$$

If $U_a(t + \Delta t - L_a/\gamma_a) - V_a(t) > c_a \cdot \Delta t$ when $\Delta t \to 0$, then the link is not in free flow condition at time t. It means that $U_a(t - L_a/\gamma_a) - V_a(t) > 0$ and

$$\lim_{\Delta t \to 0} \frac{S_a(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{c_a \cdot \Delta t}{\Delta t} = c_a$$

If $U_a(t + \Delta t - L_a/\gamma_a) - V_a(t) \leq c_a \cdot \Delta t$ when $\Delta t \to 0$, then the link is in free flow condition at time t. It means that $U_a(t - L_a/\gamma_a) - V_a(t) = 0$ and

$$\lim_{\Delta t \to 0} \frac{S_a(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{U_a\left(t + \Delta t - \frac{L_a}{\gamma_a}\right) - V_a(t)}{\Delta t}$$
$$= \frac{U_a\left(t + \Delta t - \frac{L_a}{\gamma_a}\right) - U_a\left(t - \frac{L_a}{\gamma_a}\right)}{\Delta t} = u_a\left(t - \frac{L_a}{\gamma_a}\right)$$

Similarly, the receiving flow rate is obtained by dividing the receiving flow by Δt and letting $\Delta t \to 0$:

$$r_a(t) = \lim_{\Delta t \to 0} \frac{R_a(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\min\left(V_a\left(t + \Delta t + \frac{L_a}{w_a}\right) - U_a(t) + K_a L_a; c_a \cdot \Delta t\right)}{\Delta t}$$

If $V_a(t + \Delta t + L_a/w_a) - U_a(t) + K_a L_a > c_a \cdot \Delta t$ when $\Delta t \to 0$ then the link is not in congested condition at time t. It means that $V_a(t + L_a/w_a) - U_a(t) + K_a L_a > 0$ and

$$\lim_{\Delta t \to 0} \frac{R_a(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{c_a \cdot \Delta t}{\Delta t} = c_a$$

If $V_a(t+\Delta t+L_a/w_a)-U_a(t)+K_aL_a \leq c_a \cdot \Delta t$ when $\Delta t \to 0$ then the link is in congested condition at time t. It means that $V_a(t+L_a/w_a)-U_a(t)+K_aL_a=0$ and

$$\lim_{\Delta t\to 0}\frac{R_a(t)}{\Delta t}=\lim_{\Delta t\to 0}\frac{V_a\left(t+\Delta t+\frac{L_a}{w_a}\right)-U_a(t)+K_aL_a}{\Delta t}$$
$$=\lim_{\Delta t\to 0}\frac{V_a\left(t+\Delta t+\frac{L_a}{w_a}\right)-V_a\left(t+\frac{L_a}{w_a}\right)}{\Delta t}=v_a\left(t+\frac{L_a}{w_a}\right)$$

Summarizing:

$$s_a(t)=\begin{cases}u_a\left(t-\frac{L_a}{\gamma_a}\right) & if\ U_a\left(t-\frac{L_a}{\gamma_a}\right)-V_a(t)=0\\ c_a & if\ U_a\left(t-\frac{L_a}{\gamma_a}\right)-V_a(t)>0\end{cases} \quad (21)$$

$$r_a(t)=\begin{cases}v_a\left(t+\frac{L_a}{w_a}\right) & if\ V_a\left(t+\frac{L_a}{w_a}\right)-U_a(t)+K_aL_a=0\\ c_a & if\ V_a\left(t+\frac{L_a}{w_a}\right)-U_a(t)+K_aL_a>0\end{cases} \quad (22)$$

The derivation above can also intuitively be found. If the downstream end of the link is not congested (equation (9) holds: $V_a(t)=U_a(t-L_a/\gamma_a)$), the highest possible outflow rate $s_a(t)$ is equal to the outflow rate $v_a(t)$ which is in that case equal to $u_a(t-L_a/\gamma_a)$ (see (19)). A higher outflow rate is not possible, since there are only $v_a(t)$ vehicles per hour (driving at free flow speed) available to flow out. If there is congestion at the end of the link, the sending flow rate equals the capacity. This can be explained as follows. If there is congestion at the end of the link, vehicles drive slower and closer to each other than in free flow conditions. Would the capacity constraints at the downstream link end be neglected, then vehicles could drive out of the link faster, as fast as free flow speed, meaning a flow rate equal to the link capacity. This would correspond to a transition at the fundamental diagram (Figure 5a) from somewhere at the congested part to the top of the diagram.
If the upstream end of the link is congested (equation (10) holds: $U_a(t)=V_a(t+L_a/w_a)+K_aL_a$), the highest possible inflow rate $r_a(t)$ is equal to the inflow rate $u_a(t)$ which is in that case equal to $v_a(t+L_a/w_a)$ (see (20)). Vehicles cannot enter the link at a higher rate, because there is a queue. If there is no congestion, the highest possible inflow rate equals the capacity.

The same node model as in the original LTM is used (equation (17)), but sending, receiving and transition flow rates are used instead of sending, receiving and transition flows.

$$g_{ab}(t)=\min_{\substack{b'\in E,\\ a\to b'}}\left(\frac{r_{b'}(t)}{\sum_{\substack{a'\in E,\\ a'\to b}}s_{a'b'}(t)}\cdot s_{ab}(t);\ s_{ab}(t)\right) \qquad \forall a,b\in E; a\to b \quad (23)$$

$$u_b(t) = \sum_{\substack{a' \in E, \\ a' \to b}} g_{a'b}(t) \qquad \forall b \in E \qquad (24)$$

$$v_a(t) = \sum_{\substack{b' \in E, \\ a \to b'}} g_{ab'}(t) \qquad \forall a \in E \qquad (25)$$

To determine the directional sending flow rates $s_{ab}(t)$, the turn fractions are used. The turn fraction tf_{ab} between consecutive links a and b is the fraction of traffic at link a that wants to travel towards link b. The sum of turn fractions from a link equals one, since no traffic is lost. In this thesis, it is assumed that the turn fractions remain constant during the calculations.

$$tf_{ab} = \sum_{\substack{p \in \mathcal{P}; \\ a,b \in p}} f_p \Big/ \sum_{\substack{p \in \mathcal{P}; \\ a \in p}} f_p \qquad \forall a, b \in E; a \to b \qquad (26)$$

The assumption that the turn fractions are constant during the whole simulation period can cause incorrect directional sending flows and therefore also incorrect transition flows. For instance, consider two paths that go via link a, but have a different link before and after link a on the path. If the distance that the flow has to cover before reaching link a is large for the first path and small for the second, there will be a period that only flow from the second path is on the link, and the sending flow would consist only of flow going to the next link of the second path. So, the turn fraction to the next link on the second path should be 1, and not the constant value determined by the total path flows via equation (26).
However, it is complex to keep track of the turn fractions at any time. It is not always clear which fractions of traffic wants to go each next link. Therefore, it is chosen to fix the turn fractions. More research should be done to investigate the influence of fixed turn fractions, and how variable turn fractions could be implemented.

The model also needs to keep track of the cumulative in- and outflows for each t: $U_a(t)$ and $V_a(t)$.

$$U_a(t) = \int_0^t u_a(s) ds \qquad \forall a \in E \qquad (27)$$

$$V_a(t) = \int_0^t v_a(s) ds \qquad \forall a \in E \qquad (28)$$

The equations for the link model ((21)-(22) and (27)-(28)) combined with those for the node model ((23)-(25)) together form the complete model. These equations have to hold for all t. In the next section, an algorithm for the model is given to obtain the solution.

3.8. Algorithm for the stationary Link Transmission Model

Before the algorithm for the stationary link transmission model is described, first two important components of the algorithm are discussed: the determination of the initial situation, and the finding of the next event. The first component is only used once, at initialization. The second component is used multiple times during the algorithm.

3.8.1. Initialization

To achieve the initial values of the variables, $t = 0$ is filled in into equations (21) and (22). It is assumed here that the cumulative in- and outflows and also the in- and outflow rates at a time $t < 0$ are zero. Then the sending flow is equal to zero for all links, since $U_a(0 - L_a/\gamma_a) - V_a(0) = 0$, so $s_a(0) = u_a(0 - L_a/\gamma_a) = 0$. This makes sense since within the LTM at $t = 0$ the network is empty and vehicles start to flow from the upstream end of the first link on each path. It will take some time (L_a/γ_a time units) until the first vehicle reaches the downstream end of the first link, such that the sending flow is greater than zero.
The receiving flow at $t = 0$ is always equal to the capacity, since $V_a(0 + L_a/w_a) - U_a(0) + K_a L_a = 0 - 0 + K_a L_a > 0$. So, $r_a(0) = c_a$.
The inflow rate of the first link on each path p at $t = 0$ is equal to the assigned flow f_p of that path. All other inflow and outflow rates are zero, since no vehicles are on the network yet. Summarizing:

$$s_a(0) = 0 \qquad \forall a \in E$$

$$r_a(0) = c_a \qquad \forall a \in E$$

$$u_a(0) = \sum_{\substack{p \in \mathcal{P}, \\ a\ first\ link\ of\ p}} f_p \qquad \forall a \in E \qquad (29)$$

$$v_a(0) = 0 \qquad \forall a \in E$$

Equation (29) determines the inflow rate of a link by summing the path flows of all the paths that have that link as the first link of the path.

3.8.2. Finding the next event

To propagate the flows through the network, an event-based approach is used. Between two events all sending and receiving flow rates and all in- and outflow rates are constant because of the stationary demand. This means the function of the cumulative in- and outflow is piecewise linear, and the in- and outflow rates are piecewise constant. Therefore only the next event on the network is relevant; the first moment that there is a change somewhere in the network in the sending or receiving flow rate. Suppose now is time t. The first oncoming change in the sending flow rate on a link a corresponds to the first moment after t that equation (9) on page 26 is true for link a. So the purpose is to find a candidate event time $t_a^{fw} > t$ such that:

$$V_a(t_a^{fw}) = U_a\left(t_a^{fw} - \frac{L_a}{\gamma_a}\right) \tag{30}$$

Here the fw in t_a^{fw} stands for forward, since the change in the sending flow rate is caused by an (implicit) forward shock wave in this link that has reached the downstream link end.
The inflow rate $u_a(t)$ and the outflow rate $v_a(t)$ are constant as long the event has not happened, so:

$$U_a(t+h) = U_a(t) + u_a(t) \cdot h$$

$$V_a(t+h) = V_a(t) + v_a(t) \cdot h$$

Now equation (30) can be rewritten as follows:

$$V_a(t) + v_a(t) \cdot \left(t_a^{fw} - t\right) = U_a(t) + u_a(t) \cdot \left(t_a^{fw} - \frac{L_a}{\gamma_a} - t\right)$$

$$\left(v_a(t) - u_a(t)\right) \cdot \left(t_a^{fw} - t\right) = -V_a(t) + U_a(t) - u_a(t) \cdot \frac{L_a}{\gamma_a}$$

So, the candidate event time for the downstream end of link a is:

$$t_a^{fw} = t + \frac{-V_a(t) + U_a(t) - u_a(t) \cdot \frac{L_a}{\gamma_a}}{v_a(t) - u_a(t)} \tag{31}$$

Similarly, the first oncoming change in the receiving flow rate corresponds to the first moment after t that equation (10) on page 27 is true. A candidate event time $t_a^{bw} > t$ must be found such that:

$$U_a(t_a^{bw}) = V_a\left(t_a^{bw} + \frac{L_a}{w_a}\right) + K_a L_a \tag{32}$$

Here the bw in t_a^{bw} stands for backward, since the change in the receiving flow rate is caused by an (implicit) backward shock wave in this link that has reached the upstream link end. Equation (32) can be rewritten as follows:

$$U_a(t) + u_a(t) \cdot (t_a^{bw} - t) = V_a(t) + v_a(t) \cdot \left(t_a^{bw} + \frac{L_a}{w_a} - t\right) + K_a L_a$$

$$\left(u_a(t) - v_a(t)\right) \cdot (t_a^{bw} - t) = -U_a(t) + V_a(t) + v_a(t) \cdot \frac{L_a}{w_a} + K_a L_a$$

So, the candidate event time for the upstream link end is:

$$t_a^{bw} = t + \frac{-U_a(t) + V_a(t) + v_a(t) \cdot \frac{L_a}{w_a} + K_a L_a}{u_a(t) - v_a(t)} \tag{33}$$

It is possible that for a certain link, at a certain time there is no valid candidate event time. When the in- and outflow rates are equal, the situation is stable and there are no shock waves on the link. The equations (31) and (33) will not yield a result because of the division by zero. But it is also possible that a calculated candidate event time is not greater than t. Then the candidate is rejected. The minimum over all remaining candidate event times in the network is the next event time:

$$\bar{t} = \min_{a \in E}\left(t_a^{fw}; t_a^{bw}\right) \tag{34}$$

In the algorithm all flow rates and cumulative number of vehicles are updated to time $\bar{t}$. The algorithm is presented in the next subsection.

3.8.3. Algorithm

A simple algorithm as the following can be used to compute the flow propagation for the stationary link transmission model. In the first step, the initial values are set, according to section 3.8.1. In step 2 the time and location of the event that will happen next is determined (see section 3.8.2). If this event time is after $endtime$, then the algorithm stops, else step 3 is performed. In step 3 the variables are updated to the event time. For each link the cumulative in- and outflows are updated, and depending on whether the event happened at a downstream or upstream link end, also the sending or receiving flow rate is changed at that location. In step 4 the transition flow rates and the in- and outflow rates are updated by the node model. Then the current time is updated and the algorithm goes back to step 2.

1. (Initialize) Set the initial values.

   $t = 0$

   $endtime = 1$

   $endtime$ is the length of the period for which the algorithm determines the flow rates on the links in the network in hours, for example 1 hour.

   For all links $a \in E$ set the initial values following section 3.8.1:

   $$s_a(t) = 0$$

   $$s_{ab}(t) = 0 \qquad \forall b \in E: a \to b$$

   $$r_a(t) = c_a$$

   $$v_a(t) = 0$$

   $$u_a(t) = \sum_{\substack{p \in \mathcal{P}, \\ a\ first\ link\ of\ p}} f_p$$

   $$U_a(t) = 0$$

$$V_a(t) = 0$$

2. (Find next event) Determine the next event time $\bar{t}$ on the network using equation (31), (33) and (34) in section 3.8.2. Assume the event happens on link $\bar{a}$, at node $\bar{n}$.

   If $\bar{t} > endtime$, go to step 6.

3. (Update) Calculate the cumulative flows at $\bar{t}$ for all links $a \in E$. Since the in- and outflow rates are constant between two events a linear increase can be done:

$$U_a(\bar{t}) = U_a(t) + u_a(t) \cdot (\bar{t} - t)$$

$$V_a(\bar{t}) = V_a(t) + v_a(t) \cdot (\bar{t} - t)$$

   Furthermore, the sending and receiving flow rates are updated. At the link end where the event happened the sending or receiving flow rate is changed, at other link ends there is no change.

   If the event happened at the downstream link end, i.e. $\bar{a} \in I_{\bar{n}}$:

$$s_{\bar{a}}(\bar{t}) = u_{\bar{a}}(\bar{t} - L_{\bar{a}}/\gamma_{\bar{a}})$$

   using equation (21). This inflow rate is available since between two events inflow rates are constant.

$$s_{\bar{a}b}(\bar{t}) = tf_{\bar{a}b} \cdot s_{\bar{a}}(\bar{t}) \qquad \forall b \in E: \bar{a} \to b$$

   Where tf_{ab} is the turn fraction, the fraction of the traffic on link a that is going to link b, which is assumed to be constant during the whole period (see also the end of section 3.7).

   If the event happened at the upstream link end, i.e. $\bar{a} \in J_{\bar{n}}$:

$$r_{\bar{a}}(\bar{t}) = v_{\bar{a}}(\bar{t} + L_{\bar{a}}/w_{\bar{a}})$$

   using equation (22). Like the inflow rate, the outflow rate is also constant between events.

   At all other link ends the sending and receiving flow rates remain unchanged:

$$r_a(\bar{t}) = r_a(t) \qquad \forall a \in E \backslash \bar{a}$$

$$s_a(\bar{t}) = s_a(t) \qquad \forall a \in E \backslash \bar{a}$$

$$s_{ab}(\bar{t}) = s_{ab}(t) \qquad \forall a \in E \backslash \bar{a}, \forall b \in E, a \to b$$

4. (Node model) Update the transition flows and in- and outflows at node $\overline{n}$ where the event happened using the node model formulation.

$$g_{ab}(\overline{t}) = \min_{\substack{b' \in E, \\ a \to b'}} \left(\frac{r_{b'}(\overline{t})}{\sum_{\substack{a' \in E, \\ a' \to b}} s_{a'b'}(\overline{t})} \cdot s_{ab}(\overline{t});\ s_{ab}(\overline{t}) \right) \qquad \forall a, b \in E: a \in I_{\overline{n}}, b \in J_{\overline{n}}$$

$$v_a(\overline{t}) = \sum_{\substack{b' \in E, \\ a \to b'}} g_{ab'}(\overline{t}) \qquad \forall a \in E: a \in I_{\overline{n}}$$

$$u_b(\overline{t}) = \sum_{\substack{a' \in E, \\ a' \to b}} g_{a'b}(\overline{t}) \qquad \forall b \in E: b \in J_{\overline{n}}$$

5. $t = \overline{t}$ and go to step 2.

6. (Final situation) Determine the final values of the cumulative in- and outflows.

$$U_a(endtime) = U_a(t) + u_a(t) \cdot (endtime - t)$$

$$V_a(endtime) = V_a(t) + v_a(t) \cdot (endtime - t)$$

Note: The above algorithm for the stationary Link Transmission Model could be implemented more efficiently. For example, only at the link end where the event happened there will be changes in the in- and outflow rates. At all the other locations there is no need for an update, since the flow rates remain the same. Also, the calculation of the travel times are not stated explicitly. Travel times can be derived from the piecewise linear functions of the cumulative in- and outflow. However, since the purpose of the stationary LTM is only to be an intermediate step between the original LTM and our Static Traffic Assignment with Queuing (STAQ) algorithm, the improvements are performed within the STAQ algorithm, which will be discussed in the next chapter. There also the travel time calculations are discussed.

3.9. Conclusion

In this chapter, the Link Transmission Model by Yperman (2007) is described. Furthermore, the theory about the fundamental diagram and kinematic waves and shock waves is presented shortly. It is shown in a lemma that calculating the shock wave arrival time explicitly yields the same result as with Newell's simplified theory of kinematic waves. This means that it is no longer needed to explicitly keep track of all the shock waves in the network. After that, the stationary variant of the LTM is derived, by assuming stationary traffic demand. Using an event based approach, less computations are needed. A simple algorithm is presented to solve the stationary link transmission model.

4. STAQ – Static Traffic Assignment with Queuing

In this chapter the Static Traffic Assignment with Queuing (STAQ) model is presented, the model currently under development at Goudappel Coffeng B.V. STAQ is a quasi-dynamic traffic assignment model, that propagates traffic flow through the network in a realistic way, consistent with kinematic wave theory. It is also able to handle spillback. The model is based on the stationary Link Transmission Model (LTM) from section 3.7 which itself is based on the LTM by Yperman (2007).

4.1. Introduction

The STAQ model is based on the stationary Link Transmission Model that was presented in section 3.7. However, some assumptions are made that are different from the stationary LTM. The main difference is the initial situation. Within the LTM it is assumed that the network is empty at the beginning $(t = 0)$, and that at that time the assigned path flows starts to flow, starting at the first link of each path. However, in the STAQ model, it is assumed that there is already a traffic flow rate everywhere on each link of every path on the network. This assumption is called instantaneous travel flow propagation, which means that travelling over a link does not take time. A vehicle reaches the other end of a link at once. In other words, the vehicles are on every link at the same time.

The reason that this assumption is used is that a lot of events can be skipped by assuming that there is already a traffic flow present everywhere on each link of a path. For example, consider a very small network with three consecutive links and one path. The stationary LTM will propagate the first vehicle at time $t = 0$ from the start of link 1. It will take L_1/γ_1 time units to reach the end of link 1, so the next event is at time L_1/γ_1. Two other events will occur when the first vehicle reaches the end of link 2, and later the end of link 3. Basically what these three events do is 'fill the network'. In a general network, with many OD pairs, links and paths, this amount of events can be very large. All these events needs to be handled by the algorithm. All these events can be skipped by filling the network in an initialization phase.
Another important reason is that in a real life situation, the network will seldom be empty, especially not during peak hours, which are the most interesting hours to analyze since then the most congestion occurs. Therefore, STAQ starts with an already filled network. Obviously, since STAQ starts with a different initial situation, the results will differ from the stationary LTM.

In the initialization phase or t_0 phase, this initial situation is determined. It is ensured that whenever the link capacity is lower than the total path flow rate that wants to go over that link, the link capacity is distributed over the paths proportionally to its flow, by reducing the outflow rates of these paths with an equal factor. So, at much path flow as possible is sent over the paths, as long as this does not violate any capacity constraints. No queues are built up yet.
At the queuing phase or t_1 phase, the vehicles start to flow through the network, and queues start to build up when the outflow rate is smaller than the inflow rate. After a certain period, for instance one hour, the final situation is determined and the travel times can be derived from the cumulative in- and outflows.

Similar to the LTM, it is assumed that the following input data is available:

- Network (V, E) including for each link $a \in E$:
  - $c_a > 0$ link capacity (vehicles/hour)
  - $L_a > 0$ link length (km)
  - $K_a > 0$ jam density, density at which flow is no longer possible (vehicles/km)
  - $\gamma_a > 0$ forward (free flow) wave speed (km/hour)
  - $w_a < 0$ backward (congested) wave speed (km/hour)

- The fundamental diagram of link a, determined by the parameters c_a, K_a and γ_a (see figure Figure 5a). w_a follows from these parameters by $w_a = c_a/(K_a - c_a/\gamma_a)$.

- $\mathcal{P}$: Set of used routes (or paths)

- $f_p > 0$: Path flow rates for each path $p \in \mathcal{P}$, the traffic flow per hour assigned to that path (vehicles/hour). It is assumed that the traffic demand is stationary, and so also the path flow rates are stationary. Path flow rates are assumed to be given, achieved by some route choice model.

The rest of this chapter is structured as follows. In section 4.2 the initialization phase is described. The formulation of this phase is presented and an algorithm to solve this phase is given. Section 4.3 is about the queuing phase of STAQ. At first it is showed how a quick approximation of the delay can be determined by assuming that spillback does not occur. After that the STAQ model with spillback effects is described, and an efficient algorithm is given. In section 4.4 the determination and meaning of the travel times is discussed.

4.2. t_0 phase

In the t_0 phase, or initialization phase, the initial situation of the network is determined. In this phase the amount of traffic that can flow in each link is determined, taking into account the link capacities, but no queues are formed yet. This phase is also called the ‘squeezing phase’, since as much of the assigned path flows as possible is squeezed from each origin towards its destination. If a link has insufficient capacity for the sum of the path flows going through that link, the outflow rates are lowered proportionally. Furthermore, the flow rates of the downstream links from a bottleneck are adjusted if necessary.

The idea of the t_0 phase is illustrated with the following example. Consider the network of Figure 14, with all links having a capacity of 2000 vehicles per hour. The path flow rate on route $a - c$ is 1500 and on $b - c$ it is 1000. Link c can only handle 2000 vehicles per hour, so the outflow rate of link a and b is lowered with factor 2000/2500 = 4/5. Then the outflow rate of link a is 1200 and of link b is 800. Note that only the outflow rate of the incoming links is reduced, in the rest of link there is still the original path flow rate present. The in- and outflow rates of link c are reduced to 2000.

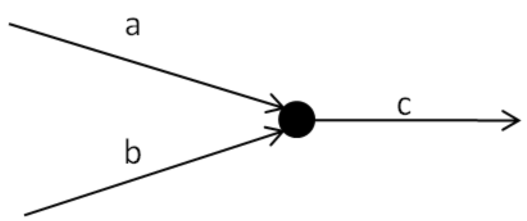


Figure 14. Small network with a merge intersection.

In the t_1 phase, queues will start to build up at the end of links a and b, against the direction of the flow, since the outflow rate is smaller than the inflow rate.

4.2.1. Formulation of the t_0 phase

The variables used in the t_0 phase are derived from the stationary LTM, however, the variables are not dependent of t, since time is not a factor in the t_0 phase.

s_a — The sending flow rate (in vehicles/hour) of link a (maximal possible outflow rate of this link when there are no restrictions downstream).

s_{ab} — The directional sending flow rate (in vehicles/hour) from link a to link b, the part of s_a that wants to go to link b. It is assumed that link a is connected to b: $a \rightarrow b$, and:

$$s_a = \sum_{\substack{b \in E, \\ a \rightarrow b}} s_{ab} \qquad \forall a \in E$$

g_{ab} — The transition flow rate (in vehicles/hour) of link a to link b (actual possible flow rate between a and b). It is assumed that $a \rightarrow b$.

u_a — The actual inflow rate (in vehicles/hour) of link a.

u_{ap} — The actual inflow rate (in vehicles/hour) of link a belonging to path p, with

$$u_a = \sum_{\substack{p \in \mathcal{P}, \\ a \in p}} u_{ap} \qquad \forall a \in E$$

v_a — The actual outflow rate (in vehicles/hour) of link a.

v_{ap} — The actual outflow rate (in vehicles/hour) of link a belonging to path p, with

$$v_a = \sum_{\substack{p \in \mathcal{P}, \\ a \in p}} v_{ap} \qquad \forall a \in E$$

The idea is to set the inflow rate on each first link of a path equal to the assigned flow rate of that path f_p, and then to force the other flow rates into the unique solution by setting constraints on the nodes and links. It is assumed that the first links of each path have enough capacity to deal with the assigned flow rate.

The t_0 problem is described by the following system of equations with unknown variables u_{ap}, v_{ap}, s_{ab} and g_{ab}, that has to be solved:

$$u_{ap} = \begin{cases} f_p & ; \text{ } if \text{ } a \text{ is the first link on path } p \\ v_{a'p} & ; \text{ } otherwise \end{cases} \quad \begin{matrix} \forall a \in E; p \in \mathcal{P}; \\ a' \in E; a', a \in p \text{ and } a' \rightarrow a \end{matrix} \tag{35}$$

$$s_{ab} = \sum_{\substack{p:\, b \in p, \\ p \in \mathcal{P}}} u_{ap} \quad \forall a, b \in E; a \rightarrow b \tag{36}$$

$$g_{ab} = \min_{\substack{\mathrm{b}' \in E, \\ a \rightarrow b'}} \left(\frac{c_{b'}}{\sum_{\substack{a' \in E, \\ a' \rightarrow b}} s_{a'b'}} \cdot s_{ab};\ s_{ab} \right) \quad \forall a, b \in E; a \rightarrow b \tag{37}$$

$$v_{ap} = \begin{cases} u_{ap}; \text{ } if \text{ } a \text{ is the last link on path } p \\ u_{ap} \cdot \dfrac{g_{ab}}{s_{ab}} \text{ } ; \text{ } otherwise \end{cases} \quad \begin{matrix} \forall a \in E; p \in \mathcal{P}; \\ b \in E; a, b \in p \text{ and } a \rightarrow b \end{matrix} \tag{38}$$

If link a is the first link on path p, the path inflow rate u_{ap} is equal to the path flow. Otherwise it is equal to the path outflow rate of the previous link on path p (equation (35)).
The directional sending flow rate s_{ab} from link a to link b is equal to the sum of the path inflow rates of link a that go to link b. This is the flow rate that actually enters link a and wants to continue to link b (equation (36)).
Equation (37) states that the transition flow rate g_{ab} from link a to link b is equal to directional sending flow rate multiplied by the factor of the largest capacity problem at this node, if any.
Finally, equation (38) ensures that if link a is the last link on path p, the path outflow rate v_{ap} is equal to the path inflow rate of that link. Otherwise it is multiplied by the ratio between the transition flow rate and the directional sending flow rate, which is equal to the ratio of the largest capacity problem at this node, if any.

Note that the receiving flow rates are not used in the formulation, since the receiving flow rates are equal to the link capacities in the t_0 phase. Therefore the receiving flow rates have no influence on the transition flows, and are replaced by the link capacities.

The sending and transition flow rates can also be eliminated from the formulation by replacing equation (36), (37) and (38) by equation (39) through substitution:

$$v_{ap} = \begin{cases} u_{ap}; & if\ a \text{ is the last link on path } p \\ u_{ap} \cdot \min\limits_{\substack{b' \in E, \\ a \to b'}} \left(\dfrac{c_{b'}}{\sum_{\substack{a' \in E, \\ a' \to b}} \sum_{\substack{p' :\, b' \in p', \\ p' \in \mathcal{P}}} u_{a'p'}};\ 1 \right); & otherwise \end{cases} \quad \begin{matrix} \forall a \in E; p \in \mathcal{P}; \\ b \in E; a, b \in p \\ \text{and } a \to b \end{matrix} \qquad (39)$$

It is assumed that all outflow rates of the incoming links of the node before a bottleneck link are reduced proportionally to the respective inflow rates of those links. In other words, all outflow rates from the incoming links are multiplied with the same factor, which equals the factor of the largest capacity problem at the node (the capacity of a link divided by the sum of the directional sending flows to that link). This factor is contained in equation (37). This means that flows not going through the bottleneck link are also reduced. Moreover, even links that do not have flow going to a bottleneck link are reduced. This is basically the same node model as is used in the (stationary) LTM in chapter 3. It is possible to use an different node model, then the determination of the transition flows should be changed (equation (37)). See also the discussion about node models at the end of section 3.6 on page 32.

The system of equations (35)-(38) is in general hard to solve, since there can exist circular dependencies between the equations. For simple networks, the equations can be easily filled in, but for general networks, an algorithm is presented in the next section. The model results in the in- and outflow rates of the t_0 phase. Queues are not yet considered, it is just the starting situation for the next phase.

4.2.2. Algorithm for the t_0 'squeezing' phase of STAQ

Next follows the algorithm for the t_0 phase. The goal is to find a solution of equations (35)-(38). The algorithm loads the path flow in fractions on the network. Each time the fraction is taken as high as possible, such that no link gets more flow than capacity when this fraction of the flow is added. A link that has a link flow equal to its capacity after the addition of this fraction is no longer considered in the algorithm, because no extra flow can go through that bottleneck link.

Variables:

$\overline{A}$	The set of 'problem links'; links that have less capacity than needed.
$\overline{N}$	The set of 'problem nodes'; nodes that have a problem link as outgoing link.
$\overline{\mathcal{P}}$	The set of all paths that go via a problem node.
A^*	The set of active problem links, subset of $\overline{A}$.
θ^i	The fraction of the total path flow that is added in iteration i.
u_{ap}^i	The inflow of link a from path p after iteration i.
u_a^i	The inflow of link a after iteration i.
f_a^i	The remaining flow that wants to go through link a after iteration i.
i	The iteration counter.

Algorithm:

1. (Initialize)

In the initialization phase, all variables are given their initial values. For the paths that do not go via any problem nodes, the whole path flow is loaded on the network (equations (41)-(42)). Other paths are not yet considered. The first fraction (43) is determined by the smallest ratio between link capacity (c_a) and demand (equation (40)).

$$f_a^0 = \sum_{p \in \mathcal{P}: a \in p} f_p \qquad \forall a \in E \qquad (40)$$

$$\overline{A} = \left\{ a \in E \,|\, \frac{f_a^0}{c_a} > 1 \right\}$$

$$\overline{N} = \left\{ n \in V \,|\, \exists a \in J_n : a \in \overline{A} \right\}$$

$$\overline{\mathcal{P}} = \left\{ p \in \mathcal{P} \,|\, \exists a \in E, a \in p : \exists n \in \overline{N} : a \in J_n \right\}$$

$$u_{ap}^0 = \begin{cases} f_p & if\ p \in \mathcal{P} \backslash \overline{\mathcal{P}} \\ 0 & if\ p \in \overline{\mathcal{P}} \end{cases} \qquad \forall a \in E, p \in \mathcal{P}: a \in p \qquad (41)$$

$$u_a^0 = \sum_{p \in \mathcal{P}: a \in p} u_{ap}^0 \qquad \forall a \in E \qquad (42)$$

$$\theta^1 = \min_{a \in \overline{A}} \left\{ \frac{c_a}{f_a^0} \right\} \qquad (43)$$

$$i = 1$$

2. (Assign increment)

Step 2 is the iteration step. In iteration i, the fraction θ^i of the path flow is added to the links (equation (44)-(45)). After that, the set A^* is updated. A^* contains all links with link flow equal to the capacity since this iteration. For each of these links, equation (47) removes from all paths that go via the upstream node of this link, all the links after this problem node from these paths. For example, if path p consists of link a-b-c-d-e, and link d is an element of A^*, then after this iteration, path p consists of link a-b-c. But also the paths that do not go through link d but do go via the upstream node of link d are shortened. These adjustments are made to ensure that no more flow originating from these paths is added to these links in upcoming iterations, because the paths are blocked at that certain node, so no more flow can go through.
Then the fraction for the next iteration is set (50), based on the flow/capacity ratios of the updated set of problem links $\overline{A}$. This procedure ensures that no flow will be added to links that already reached their capacity.

$$u_{ap}^{i} = \begin{cases} u_{ap}^{i-1} + f_p \cdot \theta^i & if\ p \in \overline{\mathcal{P}}, a \in p \\ u_{ap}^{i-1} & otherwise \end{cases} \qquad \forall a \in E, p \in \mathcal{P} \tag{44}$$

$$u_a^i = \sum_{p \in \mathcal{P}} u_{ap}^i \qquad \forall a \in E \tag{45}$$

$$A^* = \left\{ a \in \overline{A} \middle| \ \frac{u_a^i}{c_a} = 1 \right\} \tag{46}$$

Remove links a until destination from path p $\quad \forall a \in E, p \in \overline{\mathcal{P}}: a \in p| \ \exists a^* \in A^*, n \in \overline{N}: a, a^* \in J_n \quad$ (47)

$$\overline{A} = \overline{A} \backslash A^* \tag{48}$$

$$f_a^i = \sum_{p \in \mathcal{P}: a \in p} f_p \qquad \forall a \in E \tag{49}$$

$$\theta^{i+1} = \begin{cases} 1 - \sum_{j=1}^{i} \theta^j & if\ \overline{A} = \emptyset \\ \min_{a \in \overline{A}} \left\{ \frac{c_a - u_a^i}{f_a^i} \right\} & otherwise \end{cases} \tag{50}$$

$$i = i + 1$$

3. (Stopping criterium)

If the sum of all added fractions equals one, no more flow needs to be added and the algorithm jumps out of the loop.

$$if \left(\sum_{j=1}^{i-1} \theta^j = 1 \right) then \text{ go to step 4; } else \text{ go to step 2}$$

4. (Final step)

The final values are set. This step results in the in- and outflow rates of all the links in the network, such that no link has a higher flow rate than the capacity. These in- and outflow rates are used as the input for the t_1 phase.

Reset all paths to the original state (i.e. undo all removals from equation (47))

$$u_{ap} = u_{ap}^i \qquad \forall a \in E, p \in \mathcal{P}: a \in p$$

$$u_a = \sum_{p \in \mathcal{P}: a \in p} u_{ap} \qquad \forall a \in E$$

$$v_{ap} = \begin{cases} u_{ap} & if\ a \text{ is the last link of } p \\ u_{bp} & otherwise;\ a \to b, b \in p \end{cases} \qquad \forall a \in E, p \in \mathcal{P}: a \in p$$

$$v_a = \sum_{p \in \mathcal{P}: a \in p} v_{ap} \qquad \forall a \in E$$

It is not proven here that the algorithm always yields a solution that solves the formulation of the t_0 phase (equations (35)-(38)). However, it is assumed that this is the case, based on numerical tests on small and larger networks.

4.3. t_1 phase

In the t_1 phase the traffic flow is propagated through the network for a certain period of time, for instance one hour. This period is called the simulation period. In this phase the actual queuing takes place. Therefore this phase is also called the queuing phase. The in- and outflows from the t_0 phase are used as input for the t_1 phase. Before the t_1 phase of the STAQ model is described, it is showed how results could be obtained quickly by ignoring spillback effects.

4.3.1. t_1 phase of STAQ without spillback

If it is assumed that spillback does not occur, then queues do not continue on preceding links when they arrive at a link end. On a link there will appear a shock wave at $t = 0$ when the t_0 inflow rate is greater than the t_0 outflow rate, because there is a change in traffic states. This shock wave will have the following speed (see equation (8) on page 24):

$$\omega = \frac{v_a - u_a}{k_v - k_u}$$

Here k_v is the density at the end of the link, which can be taken from the congested part of the fundamental diagram at outflow rate v_a (more precisely, $k_v = K_a + v_a/w_a$). Furthermore, k_u is the density at the beginning of the link, which equals u_a/γ_a, because the traffic is flowing in at free flow speed. The shock wave speed is equal to the speed that the queue is growing, since the shock wave indicates where the congestion begins. The queue length can be found by multiplying the (negative of the) shock wave speed ω with the length of the simulation period: $-\omega \cdot endtime$.

The amount of delay given a certain queue length can be calculated as follows. The speed the vehicles drive in the queue is v_a/k_v km per hour, so the time it takes to drive through the queue is

$-\omega \cdot endtime \cdot k_v/v_a$ hour. In the part of the link with no queue, vehicles drive with free flow speed. This takes $(L_a - -\omega \cdot endtime)/\gamma_a$ hour. The total delay at this link at the end of the studied period is the sum of the free flow part and the congested part, minus the free flow travel time of the whole link:

$$delay_a = -\omega \cdot endtime \cdot \frac{k_v}{v_a} + \frac{(L_a + \omega \cdot endtime)}{\gamma_a} - \frac{L_a}{\gamma_a}$$

$$= -\omega \cdot endtime \cdot \frac{k_v}{v_a} + \frac{L_a}{\gamma_a} + \frac{\omega \cdot endtime}{\gamma_a} - \frac{L_a}{\gamma_a}$$

$$= \omega \cdot endtime \cdot \left(\frac{1}{\gamma_a} - \frac{k_v}{v_a}\right)$$

Note that this formulation is independent of the link length, which is logical since the delay is the extra time it takes to travel through the queue. But because it is assumed that spillback does not occur, queues can grow larger than the link length. Therefore the queues should not be interpreted as horizontal queues, but as a vertical queue at the downstream link end. The resulting delays can be used as a fast and easy first approximation of the amount of delay that will occur in the network given certain in- and outflows from the t_0 phase.

4.3.2. t_1 phase of STAQ with spillback

For a more realistic result the model needs to account for spillback. The following formulation is used, which is similar to the formulation of the stationary Link Transmission Model from section 3.7. The same derivation is used to achieve the formulations for the sending and receiving flow rates, and the same node model is used.

If the downstream end of a link is not congested (equation (9) holds: $V_a(t) = U_a(t - L_a/\gamma_a)$), the highest possible outflow rate $s_a(t)$ is equal to the outflow rate $v_a(t)$ which is in that case equal to $u_a(t - L_a/\gamma_a)$ (see (19)). If there is congestion at the end of the link, the sending flow rate equals the capacity. So, for all t the following equation holds:

$$s_a(t) = \begin{cases} u_a\left(t - \frac{L_a}{\gamma_a}\right) & if\ U_a\left(t - \frac{L_a}{\gamma_a}\right) - V_a(t) = 0 \\ c_a & if\ U_a\left(t - \frac{L_a}{\gamma_a}\right) - V_a(t) > 0 \end{cases} \quad (51)$$

If the upstream end of the link is congested (equation (10) holds: $U_a(t) = V_a(t + L_a/w_a) + K_a L_a$), the highest possible inflow rate $r_a(t)$ is equal to the inflow rate $u_a(t)$ which is in that case equal to $v_a\left(t + \frac{L_a}{w_a}\right)$ (see (20)). If there is no congestion, the highest possible inflow rate equals the capacity. So, for all t the following equation holds:

$$r_a(t) = \begin{cases} v_a\left(t + \frac{L_a}{w_a}\right) & if\ V_a\left(t + \frac{L_a}{w_a}\right) - U_a(t) + K_a L_a = 0 \\ c_a & if\ V_a\left(t + \frac{L_a}{w_a}\right) - U_a(t) + K_a L_a > 0 \end{cases}$$

The directional sending flow rates $s_{ab}(t)$ are determined through the turn fractions. Turn fraction tf_{ab} is the fraction of traffic in link a that wants to go to link b. It is assumed that the turn fractions remain constant during the complete t_1 phase. The turn fractions are based on the incoming flows from the t_0 phase.

$$tf_{ab} = \left(\sum_{\substack{p \in \mathcal{P}; \\ b \in p}} u_{ap}^{(t_0)} \right) / u_a^{(t_0)} \qquad \forall a, b \in E; a \to b$$

As in the stationary LTM, the assumption of constant turn fractions can cause incorrect directional sending flows. However, the influence of this flaw will be lower as within the stationary LTM, because of the assumption of instantaneous travel flow propagation. The t_1 phase starts from a network with a traffic flow rate on each link, based on the total path flow that wants to go over that link, independent of the length of the path preceding that link.
The sending and receiving flow rates together form the link model. The following equations, that must hold for all t, form the node model:

$$g_{ab}(t) = \min_{\substack{b' \in E, \\ a \to b'}} \left(\frac{r_{b'}(t)}{\sum_{\substack{a' \in E, \\ a' \to b}} s_{a'b'}(t)} \cdot s_{ab}(t);\ s_{ab}(t) \right) \qquad \forall a, b \in E; a \to b \qquad (52)$$

$$u_b(t) = \sum_{\substack{a' \in E, \\ a' \to b}} g_{a'b}(t) \qquad \forall b \in E$$

$$v_a(t) = \sum_{\substack{b' \in E, \\ a \to b'}} g_{ab'}(t) \qquad \forall a \in E$$

$$v_{ap}(t) = u_{bp}(t) \qquad \forall a, b \in E; p \in \mathcal{P};\ a, b \in p \text{ and } a \to b$$

This is the same node model as in the stationary LTM. For a discussion of this node model see the end of section 3.6 on page 32.
The above equations for the link and node model are the same as the equations for the stationary Link Transmission Model. The difference between both models is the situation it starts from. The stationary LTM starts from an empty network, and only at the first link of each path traffic flow is entering the network. Within STAQ, it is assumed that there is instantaneous travel flow propagation, such that everywhere on the link is a flow rate. The t_0 phase is used to find the initial in- and

outflows. However, the formulation of the LTM is based on the relation between the cumulative number of vehicles that have entered and have left a link. Because the in- and outflows of the links are set through the result of the t_0 phase instead of starting with an empty network, these formulations cannot be used directly.
It is assumed that at the start there is a traffic flow rate equal to the inflow rate everywhere on the link, in free flow conditions. This could be seen as a number of vehicles that already have entered the link at $t = 0$ equal to the inflow rate times the free flow travel time:

$$U_a(0) = u_a \cdot \frac{L_a}{\gamma_a}$$

In this way, the relation between the number of vehicles that entered the link and number of vehicles that left the link is consistent again, and the formulations from the stationary Link Transmission Model can be used.

4.3.3. Algorithm for STAQ t_1 phase

Now the algorithm for the t_1 phase is presented. The following variables are used:

s_a	The sending flow rate of link a.
s_{ab}	The directional sending flow rate of link a to link b.
r_a	The receiving flow rate of link a.
$events_a^u$	Array with the times of the events that happened at the node at the upstream end of link a.
$events_a^v$	Array with the times of the events that happened at the node at the downstream end of link a.
u_a	Array with the inflow rate of link a, at the times that an event happened at the node at the upstream end of link a. An inflow rate is valid until the next event time, a next event implies a new inflow rate.
v_a	Array with the outflow rate of link a, at the times that an event happened at the node at the downstream end of link a. An outflow rate is valid until the next event time, a next event implies a new outflow rate.
U_a	Array with the cumulative number of vehicles that have entered link a, at the times that an event happened at the node at the upstream end of link a.
V_a	Array with the cumulative number of vehicles that have left link a, at the times that an event happened at the node at the downstream end of link a.
t_a^{fw}	The time of the first oncoming event at the downstream end of link a provided that no other events happen on this link.

t_a^{bw} The time of the first oncoming event at the upstream end of link a provided that no other events happen on this link.

Furthermore the following sets and array operations are used:

I_n The set of incoming links into node n, i.e. all links that are connected to node n with the downstream link end.

J_n The set of outgoing links from node n, i.e. all links that are connected to node n with the upstream link end.

$name(i)$ Addresses the i^{th} element in the array $name$, $0 \leq i \leq end$.

$name(end+1)$ Adds a new element at the end of the array $name$ and addresses that element

$size(name)$ Returns the size of the array $name$.

$find(name, val)$ Returns the index of the value val in the array $name$. If $name$ does not contain val, it returns the index of the greatest value smaller than val.

The determination of the next event times is slightly different than in section 3.8.2. The event times at the downstream link end are different than the event times at the upstream end of a link, because only at the link end the event happens an event time is added to the array. Therefore the current time t cannot be used (except for the first event times). For the next event at the downstream link end t_a^{fw} needs to be found such that $V_a\left(t_a^{fw}\right) = U_a\left(t_a^{fw} - L_a/\gamma_a\right)$ holds. So:

$$V_a(end) + v_a(end) \cdot \left(t_a^{fw} - events_a^v(end)\right) = U_a(end) + u_a(end) \cdot \left(t_a^{fw} - \frac{L_a}{\gamma_a} - events_a^u(end)\right)$$

$$t_a^{fw}\left(-u_a(end) + v_a(end)\right) = U_a(end) - u_a(end) \cdot \frac{L_a}{\gamma_a} - u_a(end) \cdot events_a^u(end) - V_a(end) + v_a(end) \cdot events_a^v(end)$$

$$t_a^{fw} = \frac{\begin{pmatrix} U_a(end) - V_a(end) - u_a(end) \cdot events_a^u(end) \\ +v_a(end) \cdot events_a^v(end) - u_a(end) \cdot \frac{L_a}{\gamma_a} \end{pmatrix}}{-u_a(end) + v_a(end)}$$

Similarly, for the next event at the upstream link end t_a^{bw} must be found such that $U_a\left(t_a^{bw}\right) = V_a\left(t_a^{bw} + L_a/w_a\right) + K_a L_a$ holds. So:

$$U_a(end) + u_a(end) \cdot \left(t_a^{bw} - events_a^u(end)\right)$$
$$= V_a(end) + v_a(end) \cdot \left(t_a^{bw} + \frac{L_a}{w_a} - events_a^v(end)\right) + K_a L_a$$

$$t_a^{bw}\left(u_a(end) - v_a(end)\right)$$
$$= V_a(end) + v_a(end) \cdot \frac{L_a}{w_a} - v_a(end) \cdot events_a^v(end) + K_a L_a$$
$$- U_a(end) + u_a(end) \cdot events_a^u(end)$$

$$t_a^{bw} = \frac{\begin{pmatrix} V_a(end) - U_a(end) - v_a(end) \cdot events_a^v(end) \\ + u_a(end) \cdot events_a^u(end) + v_a(end) \cdot \frac{L_a}{w_a} + K_a L_a \end{pmatrix}}{-v_a(end) + u_a(end)}$$

The algorithm consists of an initialization step, where all variables are set to their initial values. Then follows the iteration loop. In each iteration, the next event time on the network is found. In an iteration, the next event time is processed, all necessary values are updated. If the next event time is greater than the end time of the simulation period, the algorithm jumps out of the loop to the finalization step. In that step the travel times are calculated.

1. (Initialize)
   The initial in- and outflow rate are taken from the t_0 phase. The sending flow rate is equal to the inflow rate, since it is assumed that this rate is present everywhere on the link. The receiving flow rate is initially equal to the link capacity. The candidate event times and the arrays for the event times and the cumulative in- and outflows are initialized. n/a means that there is no next event time available at this time.

   $t = 0$

   $endtime = 1$

   For all links $a \in E$:

   $u_a = \left[u_a^{(t_0)}\right]$

   $v_a = \left[v_a^{(t_0)}\right]$

   $s_a = u_a^{(t_0)}$

   $s_{ab} = s_a \cdot tf_{ab} \qquad \forall b \in E; a \to b$

   $r_a = c_a$

   $events_a^u = [0]$

$$events_a^v = [0]$$

$$U_a = [u_a(end) \cdot \frac{L_a}{\gamma_a}]$$

$$V_a = [0]$$

$$t_a^{fw} = n/a$$

$$t_a^{bw} = n/a$$

2. (Find first event)
   For each link a, determine the first event time at the beginning and the end of that link using equation (31) and (33). These values are candidate event times, they are the next event provided that no other events happen at that link. If the in- and outflow rate are equal, there will be no queue growing at $t = 0$, so there are no candidate event times available for that link. If a candidate event time is smaller than zero the value is not valid, so it is rejected.

   if$\big(u_a(0) == v_a(0)\big)$ then $t_a^{bw} = t_a^{fw} = n/a$

   else

$$t_a^{fw} = \frac{U_a(0) - V_a(0) - u_a(0) \cdot \frac{L_a}{\gamma_a}}{-u_a(0) + v_a(0)}$$

   if$\big(t_a^{fw} \leq 0\big)$ then $t_a^{fw} = n/a$

$$t_a^{bw} = \frac{V_a(0) - U_a(0) + v_a(0) \cdot \frac{L_a}{w_a} + K_a L_a}{-v_a(0) + u_a(0)}$$

   if$(t_a^{bw} \leq 0)$ then $t_a^{bw} = n/a$

3. (Start of the loop - Find time and location of next event)
   The time of the next event on the network is the smallest value of all available candidate event times. If this time is greater than the predefined length of the simulation period, the algorithm jumps out of the loop.

$$\bar{t} = \min_a\big(t_a^{fw}, t_a^{bw}\big)$$

   If ($\bar{t} > endtime$ or $\bar{t} == n/a$) go to step 7.

   The event happens at link $\bar{a}$, node $\bar{n}$.

4. (Update sending and receiving flow rates)
At the link end of $\overline{a}$ that is connected to $\overline{n}$, where the event happens, the sending or receiving flow rate is updated.

If $\overline{a} \in I_{\overline{n}}$ (event at the downstream link end of $\overline{a}$)

Update the sending flow to the value of the inflow rate $L_{\overline{a}}/\gamma_{\overline{a}}$ time units ago, since the new outflow rate at the moment of the event is equal to the inflow rate $L_{\overline{a}}/\gamma_{\overline{a}}$ time units ago by equation (19).

$$s_{\overline{a}} = u_{\overline{a}}\left(find\left(events_{\overline{a}}^{u}, \overline{t} - \frac{L_{\overline{a}}}{\gamma_{\overline{a}}}\right)\right)$$

$$s_{\overline{a}b} = tf_{\overline{a}b} \cdot s_{\overline{a}} \qquad \forall b \in J_{\overline{n}}$$

If $\overline{a} \in J_{\overline{n}}$ (event at the upstream link end of $\overline{a}$)

Update the receiving flow to the value of the outflow rate $-L_{\overline{a}}/w_{\overline{a}}$ time units ago, since the new inflow rate at the moment of the event is equal to the outflow rate $-L_{\overline{a}}/w_{\overline{a}}$ time units ago by equation (20).

$$r_{\overline{a}} = v_{\overline{a}}\left(find\left(events_{\overline{a}}^{v}, \overline{t} + \frac{L_{\overline{a}}}{w_{\overline{a}}}\right)\right)$$

5. (Node model)
Determine the transition flows at node $\overline{n}$, following equation (52).

$$g_{ab} = \min_{\substack{b' \in E, \\ a \to b'}} \left(\frac{r_{b'}}{\sum_{\substack{a' \in E, \\ a' \to b}} s_{a'b'}} \cdot s_{ab};\ s_{ab} \right) \qquad \begin{array}{l} \forall a \in I_{\overline{n}},\ b \in J_{\overline{n}}; \\ a \to b \end{array}$$

6. (Update)
For all the incoming links of $\overline{n}$, the cumulative outflow, event array, outflow rate and the next event times are updated if the outflow rate is no longer equal to the sum of the transition rates from this link.

For all $a \in I_{\overline{n}}$:

if $\left(v_a(end) <> \sum_{\substack{b' \in E, \\ a \to b'}} g_{ab'} \right)$ then

$$V_a(end + 1) = V_a(end) + v_a(end) \cdot \left(\overline{t} - events_a^{v}(end) \right)$$

$$events_a^{v}(end + 1) = \overline{t}$$

$v_a(end) = \sum_{\substack{b' \in E, \\ a \to b'}} g_{ab'}$

if$\left(u_a(end) == v_a(end)\right)$ then $t_a^{bw} = t_a^{fw} = n/a$

else

$$t_a^{fw} = \frac{\begin{pmatrix} U_a(end) - V_a(end) - u_a(end) \cdot events_a^u(end) \\ +v_a(end) \cdot events_a^v(end) - u_a(end) \cdot \frac{L_a}{\gamma_a} \end{pmatrix}}{-u_a(end) + v_a(end)}$$

if$\left(t_a^{fw} \leq t\right)$ then $t_a^{fw} = n/a$

$$t_a^{bw} = \frac{\begin{pmatrix} V_a(end) - U_a(end) - v_a(end) \cdot events_a^v(end) \\ +u_a(end) \cdot events_a^u(end) + v_a(end) \cdot \frac{L_a}{w_a} + K_a \cdot L_a \end{pmatrix}}{-v_a(end) + u_a(end)}$$

if$\left(t_a^{bw} \leq t\right)$ then $t_a^{bw} = n/a$

For all the outgoing links of $\overline{n}$, the cumulative inflow, event array, inflow rate and the next event times are updated if the inflow rate is no longer equal to the sum of the transition rates to this link.

For all $b \in J_{\overline{n}}$:

if $\left(u_b(end) <> \sum_{\substack{a' \in E, \\ a' \to b}} g_{a'b}\right)$ then

$$U_b(end + 1) = U_b(end) + u_b(end) \cdot \left(\overline{t} - events_b^u(end)\right)$$

$$events_b^u(end + 1) = \overline{t}$$

$$u_b(end) = \sum_{\substack{a' \in E, \\ a' \to b}} g_{a'b}$$

if$\left(u_b(end) == v_b(end)\right)$ then $t_b^{bw} = t_b^{fw} = n/a$

else

$$t_b^{fw} = \frac{\begin{pmatrix} U_b(end) - V_b(end) - u_b(end) \cdot events_b^u(end) \\ +v_b(end) \cdot events_b^v(end) - u_b(end) \cdot \frac{L_b}{\gamma_b} \end{pmatrix}}{-u_b(end) + v_b(end)}$$

if$\left(t_b^{fw} \leq t\right)$ then $t_b^{fw} = n/a$

$$t_b^{bw} = \frac{\begin{pmatrix} V_b(end) - U_b(end) - v_b(end) \cdot events_b^v(end) \\ +u_b(end) \cdot events_b^u(end) + v_b(end) \cdot \frac{L_b}{w_b} + K_b \cdot L_b \end{pmatrix}}{-v_b(end) + u_b(end)}$$

if$\left(t_b^{bw} \leq t\right)$ then $t_b^{bw} = n/a$

The time is updated to the event time and the loop starts over at step 3.

$$t = \bar{t}$$

go to step 3.

7. (Final step)

The next event on the network is after the end time of the simulation period, so this event is not considered. The cumulative inflows are updated to their final values at $endtime$, by a linear increase. Then it is calculated at which moment the cumulative outflow has reached the level of the cumulative inflow, based on the last outflow rate. This moment will be after $endtime$ in general.

For all links $a \in E$:

$$U_a(end+1) = U_a(end) + u_a(end) \cdot \left(endtime - events_a^u(end)\right)$$

$$events_a^u(end+1) = endtime$$

$$u_a(end+1) = 0$$

$$t_a^{fw} = \frac{U_a(end) - V_a(end)}{v_a(end)} + events_a^v(end)$$

$$V_a(end+1) = U_a(end)$$

$$events_a^v(end+1) = t_a^{fw}$$

$$v_a(end+1) = 0$$

$$U_a(end+1) = U_a(end)$$

$$events_a^u(end+1) = t_a^{fw}$$

4.4. Travel times

After the simulation period is over, the link travel times are calculated. The link travel times are derived from the cumulative in- and outflows. In Figure 15 an example of a graph of the cumulative in- and outflows is shown. This graph is based on the arrays of the cumulative in- and outflows. Between two consecutive points a straight line is drawn, since the in- and outflow rates are constant between two events. To calculate the average link travel time, the difference between the areas below $U(t)$ and $V(t)$ is calculated, and this is divided by the total number of vehicles that entered (and left) this link in $endtime$ time units.

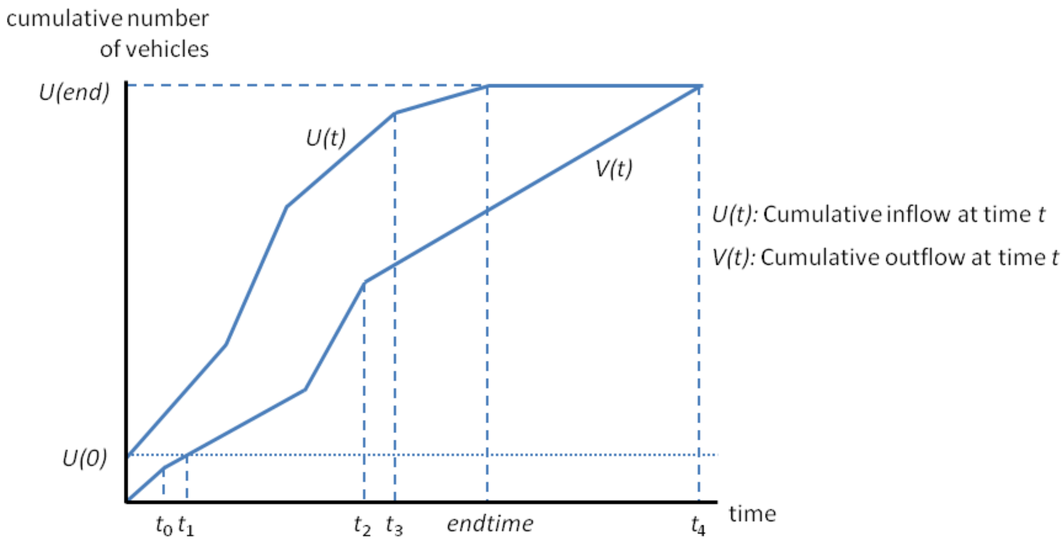


Figure 15. Derivation of the link travel time.

In this example, t_2 is the time of the last event at the downstream link end, t_3 is the time of the last event at the upstream link end. t_4 is the time that the last vehicle exits the link, i.e. t_a^{fw} in step 7 of the algorithm of the previous section.

The graph of the cumulative inflow starts at $U(0) = u \cdot L/\gamma$ vehicles, to compensate for the vehicles already present in the link (see the end of section 4.3.2). However, this amount is not included in the calculations for the average link travel time. The first vehicle that enters the link at $t = 0$, and exits the link at t_1 is the first vehicle that is in the calculations. So, the graphs of the cumulative in- and outflow can be shifted $u \cdot L/\gamma$ vehicle units down, and the part below zero is disregarded. The time axis now lies at the dotted horizontal line at $U(0)$. Then the average travel time in this hour is calculated by dividing the difference between the areas below $U(t)$ and $V(t)$ by $U(end) - U(0)$.

For all links $a \in E$:

The moment the first vehicle exits link a (t_1) is in general not an event, this moment is determined as follows:

> $index = find\big(V_a, U_a(0)\big)$, this is the index of the largest event smaller than t_1 (this can be at time $t = 0$, if no events happened before t_1).

$$\big(t_1 - events^v(index)\big) \cdot v(index) = U(0) - V(index), \text{ so}$$

$$t_1 = \frac{\big(U(0) - V(index)\big)}{v(index)} + events^v(index)$$

Then the areas under the cumulative flows are calculated, by dividing the area in $size(events_a^u) - 1$ parts and summing the area of each of those parts:

$$\begin{aligned} area_a^u &= \int_0^{events^u(end)} U(s)\, ds \\ &= \sum_{i=1}^{size(events_a^u)-1} \left(\begin{array}{c} \big(events_a^u(i) - events_a^u(i-1)\big) \cdot \\ \left(\frac{1}{2}\big(U_a(i-1) + U_a(i)\big)\right) \end{array} \right) \\ &\quad - U(0) \cdot events_a^u(end) \end{aligned}$$

The area below the dotted line in Figure 15 is subtracted by the last term. The area under the graph of the cumulative outflow is calculated in a similar way:

$$\begin{aligned} area_a^v &= \int_0^{events^v(end)} V(s)\, ds \\ &= (events_a^v(index+1) - t_1) \cdot \left(\frac{1}{2}\big(V_a(index+1) + U_a(0)\big)\right) \\ &\quad + \sum_{i=index+2}^{size(events_a^v)-1} \left(\begin{array}{c} \big(events_a^v(i) - events_a^v(i-1)\big) \cdot \\ \left(\frac{1}{2}\big(V_a(i-1) + V_a(i)\big)\right) \end{array} \right) \\ &\quad - U(0) \cdot (events_a^u(end) - t_1) \end{aligned}$$

The average link travel time is determined as follows:

$$\tau_a = \frac{area_a^u - area_a^v}{U_a(end) - U(0)}$$

The average path travel time can be found by summing the link travel times of that path, where it is assumed that the path costs are additive.

$$\tau_p = \sum_{\substack{a \in E, \\ a \in p}} \tau_a \qquad \forall p \in \mathcal{P}$$

As was shown in the traffic assignment procedure in Figure 1 on page 1, the path travel times are tested by some predefined convergence criteria, to check whether the result is acceptable. If not, the path travel times are used as input for the next iteration of the route choice model, to distribute the OD demand over the different paths in a better way, finally resulting in lower travel times.

In this thesis it is always assumed that the (stationary) route flows are given. The stationary route flows are obtained by some route choice model, based on the stationary origin-destination (OD) demand matrix. The stationary OD demand is derived from the dynamic OD demand. However, this derivation is not a trivial task. The way the average travel times of STAQ have to be interpreted depends on the interpretation of the stationary OD demand that was used.

A possible way to choose the stationary OD matrix is by taking the average demand from the dynamic matrix for each OD pair individually for a certain period. Suppose this period is one hour during the morning rush hours. The resulting stationary route flows from the route choice model can then be seen as the average traffic demand during that hour. Performing STAQ for an hour on these route flows will result in average travel times for each link during that hour in the morning rush hours. So, the average delay that a vehicle, that is travelling somewhere within that hour, will experience on a certain path is equal to the average path travel time from STAQ for this path minus the free flow path travel time.

This approach only works when in the studied period the average traffic demand does give capacity problems. If this is not the case, performing STAQ will result in free flow travel times while in the dynamic case, there may be congestion. So, STAQ is particularly useful during very congested periods.

The results from the stationary Link Transmission Model can be used to approximate the average path travel times if the network is empty at the beginning of the period that is modeled. This is because within the stationary LTM it is assumed that the network is empty at the beginning of the simulation period. The period should then be chosen at the beginning of the rush hours, because then there is not much traffic on the network yet.

The stationary demand can also be chosen based on the maximum dynamic demand in a certain period. However, the results from STAQ will then be overestimated, since then it is assumed that this maximum demand is present during the whole period instead only at a certain moment in time.

It is important to notice that the results of both the stationary LTM as the STAQ model should be interpreted as average travel times over all vehicles in a certain time period. The resulting travel time should not be interpreted as the travel time of a single vehicle. The results should only be used for traffic flows of vehicles on a macroscopic level of detail.

5. Conclusions and recommendations

In this thesis, an overview was given of the different static traffic assignment models that can deal with congestion phenomena and possibly spillback effects that are found in the literature.
The first research question:

1. Are there static traffic assignment models in the literature that can deal with congestion and can compute realistic travel times?

The different approaches and methods from literature were discussed and compared. None of the discussed static models in the literature can both deal with congestion and compute realistic travel times. The need for a model that handles congestion in a realistic way and is able to model spillback effects led to the second research question of this thesis:

2. Can a model be specified that propagates traffic correctly through the network and computes more realistic travel times than current static models, and can still compute them in a reasonable time?

To answer the second research question, it was proposed to take the Link Transmission Model by Yperman (2007) as a starting point. The LTM is a dynamic network loading model. First the working of the LTM was explained, supported by the necessary theory about kinematic waves. Next, the static variant of the LTM was derived. Traffic flows are assumed stationary, and therefore constant in time. The stationary flows can be seen as the average traffic flow during a certain period. Through this assumption, it is no longer necessary to calculate the cumulative number of vehicles for every small time step. Through an event based approach, it is possible to do this in much larger steps, such that far less computations are needed. It was proven with a lemma that the event times that are calculated with the stationary LTM are equal to the arrival times of the shock waves if they are calculated explicitly.

From the stationary Link Transmission Model the STAQ model was developed. In the STAQ – Static Traffic Assignment with Queuing – model it is assumed that the network is initially not empty, but there is instantaneous travel flow propagation, which means that there is already a traffic flow rate on the network, without any queues. This assumption is made because in reality the network will not be empty at the beginning. Besides, it decreases the number of events. The initial situation is determined in the t_0 phase, based on the given path flows. In the t_0 phase the path flows are squeezed through the network such that there is no traffic flow rate greater than the capacity in any link. A system of equations was formulated to find the unique solution. This system of equations is hard to solve mathematically, but an algorithm was provided to solve this problem.
In the second phase of STAQ, the t_1 phase, the traffic flow is propagated through the network and queues are created if necessary. Also spillback is taken into account. The flow propagation is according to kinematic wave theory, similar to the link transmission model, with an event based approach based on the cumulative number of vehicles that have entered and left each link. The algorithm results in the link and path flows during the simulation period, and more importantly, the average link travel times which can be used in the next iteration of the route choice model.

In the STAQ model, traffic flow is propagated over the links in a realistic way that is consisted with kinematic and shock wave theory. Congestion is built up and spillback is taken into account. Since other static models do not have all these properties, the STAQ model will in theory yield more realistic link travel times. However, more research is needed to prove this also empirically.
The resulting average travel times from STAQ can be seen as the average travel time of a vehicle that travels somewhere during the considered period. It is important to keep in mind how the stationary traffic demand was chosen from the dynamic demand when interpreting the results.

In this thesis, it was not possible to give each detail the attention it deserves. Therefore, some recommendations for possible improvements and further research directions for STAQ are given.

Empirical testing
The STAQ model is implemented in Matlab and tested on some small networks and one larger network. STAQ is theoretically superior to other static traffic assignment models due to the way congestion and spillback are modeled. However, to prove this claim more testing is needed on large networks to compare the results empirically with other static traffic assignment models. It should be investigated whether STAQ generates results that are closest to reality and still has an acceptable computation time.

Node model
The node model that is used in this thesis could be improved. The current model is chosen for its simplicity, the outflow rates from all directions are decreased with the same ratio if there is a capacity problem at a node. A first step would be to change the model such that incoming links that do not have flow going through a bottleneck link are not decreased.
Tampère e.a. (2011) criticize the node model that is used in this thesis. They show that the total flow is not always maximized. Furthermore, the invariance principle is not satisfied. The findings of Tampère e.a. come from a dynamic context, so it should be investigated what the impact is of these flaws of the model in a static environment, and whether the current model can be adapted to correct for this flaws, or that another node model should be implemented. If a new model is needed, the list of requirements for a node model by Tampère e.a. should be kept in mind.

t_0 algorithm
The algorithm that was presented for the t_0 phase solves the t_0 problem, but the algorithm could be implemented more efficiently. The current communication between Matlab (in which the prototype was implemented) and the transport planning software OmniTrans (which is developed and used at Goudappel Coffeng B.V.) is not very efficient. By implementing the algorithm in OmniTrans this is no longer an issue.
Another possible improvement is to introduce a minimal fraction size at an iteration. This can cause some loss in precision because some links may receive a little more flow than capacity, but it can reduce the computation time significantly. Besides, the influence of this surplus is unclear. More research is needed to this tradeoff between precision and speed.

Necessity of t_0 phase

At this moment the t_0 phase takes much more time to compute than the t_1 phase. As proposed in the recommendation above, some improvements can be made. However, it is also useful to investigate the effect of the assumption of instantaneous travel flow propagation. Maybe the t_0 phase does not add enough value to make up for the extra computation time. An alternative approach would be to introduce a different initialization phase than t_0, by running the t_1 phase for a certain period, starting with an empty network as with the stationary LTM. After this preliminary period there is a traffic flow on the network, and the t_1 phase for the studied period can be run.

Variable turn fractions

As briefly discussed in sections 3.7 and 4.3.2, the assumption of fixed turn fractions can cause incorrect transition flows. However, the size of this error is not yet clear. More research is needed to investigate whether the error is acceptable. If not, more effort should be made to implement variable turn fractions. By keeping track of all u_{ap}'s (inflow rate at link a from path p) it should be possible to do so. The directional sending flows can be derived from the incoming flow rates similar to the sending flows (see equation (51)). The turn fractions can then be derived from the directional sending flow rates, instead of the other way round.

Implementation in route choice cycle

In this thesis it is assumed that the route flows from an existing route choice model are used as the input of the STAQ model. However, it is useful to investigate how the stationary demand should be chosen from the dynamic demand over a certain period.

The results of the STAQ model (link travel times) can then be used in the next iteration of the route choice model. More research is needed to investigate how these achieved results should be used in the next iteration.

Junction modeling

Currently, STAQ handles each intersection as an unsignalized junction. However, to be more realistic, signalized or give way intersections should be modeled as such. Therefore it should be investigated whether existing junction modeling techniques can be implemented into the STAQ model.

Making STAQ more dynamic

A possible extension of the STAQ model would be to introduce a parameter which sets the amount of parts in which the studied period is divided. Each of those periods can have a different set of route flows and STAQ is performed individually. This creates a hybrid solution between a fully dynamic and a quasi-dynamic model. For instance the morning rush hour could be divided in 6 periods of 10 minutes instead of one period of one hour.

References

4Cast (2009). Qblok_2004: toedelingsprocedure LMS/NRM. (*in Dutch; working report*)

Beckmann, M., McGuire, C. & Winsten, C. (1956). *Studies in the economics of transportation*. Yale University Press, New Haven, CT.

Bifulco, G. & Crisalli, U. (1998). Stochastic user equilibrium and link capacity constraints: formulation and theoretical evidences. *Proceedings of the European Transport Conference 1998.*

Bundschuh, M., Vortisch, P. & Van Vuuren, T. (2006). Modelling queues in static traffic assignment. *Proceedings of the European Transport Conference 2006*.

Chudak, F., Eleuterio, V. & Nesterov, Y. (2007). Static traffic assignment problem. A comparison between Beckmann (1956) and Nesterov & de Palma (1998) models. *Conference paper Swiss Transport Research Conference 2007*.

Daganzo, C. (1994). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research*, Vol. 28B, pp. 269-287.

Daganzo, C. (1997). *Fundamentals of transportation and traffic operations*. Pergamon-Elsevier, Oxford.

Daganzo, C. (1998). Queue spillovers in transportation networks with a route choice. *Transportation Science*, Vol. 32(1), pp. 3-11.

Faigle, U., Kern, W. & Still, G. (2002). *Algorithmic principles of mathematical programming*. Kluwer Academic Publisher, Dordrecht.

Green, G. (1828). *An essay on the application of mathematical analysis to the theories of electricity and magnetism*. T. Wheelhouse, Nottingham.

Kern, W. & Still, G. (2009). Wardrop vs Nesterov traffic equilibrium model: Relations and differences. *Working paper*.

Larsson, T. & Patriksson, M. (1992). Simplicial decomposition with disaggregated representation for the traffic assignment problem. *Transportation Science*, Vol. 26, pp. 4-17.

Larsson, T. & Patriksson, M. (1995). An augmented lagrangean dual algorithm for link capacity constrained traffic assignment problems. *Transportation Research*, Vol. 29B, pp. 433-455.

Larsson, T. & Patriksson, M. (1999). Side constrained traffic equilibrium models – analysis, computation and applications. *Transportation Research*, Vol. 33B, pp. 233-264.

Marcotte, P., Nguyen, S. & Schoeb, A. (2004). A strategic flow model of traffic assignment in static capacitated networks. *Operations Research*, Vol. 52, No. 2, pp. 191-212.

Maugeri, D. (1994). Optimization problems with side constraints and generalized equilibrium principles. *Le Matematiche*, Vol. 49, pp. 305-312.

Nesterov, Y. (2000). Stable traffic equilibria: properties and applications. *Optimization and Engineering*, 3, 29-50.

Nesterov, Y. & de Palma, A. (2000a). Stable dynamics in transportation systems. *CORE discussion paper*.

Nesterov, Y. & de Palma, A. (2000b). Stable dynamics transportation systems. *Proceedings of the European Transport Conference 2000*.

Nesterov, Y. & de Palma, A. (2003). Stationary dynamic solutions in congested transportation networks: summary and perspectives. *Network and Spatial Economics*, Vol. 3, pp. 371-395.

Newell, G. (1993). A simplified theory of kinematic waves in highway traffic, Part I: General theory, Part II: Queuing at freeway bottlenecks, Part III: Multi-destination flows. *Transportation Research*, Vol. 27B, pp. 281-313.

Nie, Y., Zhang, H. & Lee, D. (2004). Models and algorithms for the traffic assignment problem with link capacity constraints. *Transportation Research*, Vol. 38B, pp. 285-312.

Prashker, J. & Toledo, T. (2004). A gradient projection algorithm for side-constrained traffic assignment. *European Journal of Transport and Infrastructure Research*, Vol. 4, No. 2, pp. 177-193.

Shahpar, A., Aashtiani, H. & Babazadeh, A. (2008). Dynamic penalty function method for the side constrained traffic assignment problem. *Applied Mathematics and Computation*, 206, pp. 332-345.

Smith, M. (1979). The existence, uniqueness and stability of traffic equilibria. *Transportation Research*, Vol. 13B, pp. 295–304.

Smulders, S. (1990). Control of freeway traffic flow by variable speed signs. *Transportation Research*, Vol. 24B, pp. 111-132.

Tampère, C.M.J., Corthout, R., Cattrysse, D. & Immers, L.H. (2011). A generic class of first order node models for dynamic macroscopic simulation of traffic flows. *Transportation Research*, Vol. 45B, pp. 289-309.

US Bureau of Public Roads (1964). *Traffic assignment manual*. US Government Printing Office, Washington DC.

Wardrop, J. (1952). Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers,* Part II, pp. 325–378.

Yang, H. & Huang, H. (2005). *Mathematical and economic theory of road pricing*. Elsevier, Oxford.

Yperman, I. (2007). *The link transmission model for dynamic network loading*. PhD Thesis, Faculty of Engineering, Katholieke Universiteit Leuven.