Combining static and dynamic models for traffic signal optimization

Inherent load-dependent travel times in a cyclically time-expanded network model

Ekkehard Köhler, Martin Strehler*
Brandenburgische Technische Universität, Postfach 10 13 44, 03013 Cottbus, Germany

Abstract

Travel times in traffic models are of great interest for developing realistic solutions for traffic assignment and traffic signal coordination in urban traffic networks. In this paper, we present a cyclically time-expanded network for this purpose, which is a static and linear model from its structure. However, it provides enough dynamics to reproduce load-dependent travel times and it is capable to model traffic signals. Thus, this traffic flow model is at the cutting-edge of static and dynamic models, and furthermore, it allows the simultaneous optimization of traffic assignment and signal coordination with exact mathematical programming techniques. We study the inherent properties of the travel times in this model and demonstrate its capabilities by simulation results obtained with state-of-the-art simulation tools.

© 2012 The Authors. Published by Elsevier Ltd. Selection and/or peer-review under responsibility of the Program Committee

Keywords: traffic signal optimization; traffic assignment; cyclically time-expanded network; flow over time; link performance

1. Motivation

Traffic signals can be seen as a backbone in the control of traffic flows in urban areas. Since the appearance of the first signalized intersections over 100 years ago, the improvement of signal control strategies has been an important subject of research. Friedrich (2002) and Papageorgiou et al. (2003) provide overviews of the main research lines in this area over the last years. Recently, we also presented a model for traffic signal coordination based on cyclically time-expanded networks (see Köhler & Strehler (2010)). In contrast to most of previous

* Corresponding author. Tel.: +49-355-69-2981; fax: +49-355-69-3506.
E-mail addresses: [ekoehler,strehler]@math.tu-cottbus.de
mathematical approaches, we focused not only on optimizing the coordination, but also mapped the implied traffic assignment. Each change in the signal parameters may influence the travel times in the network. Consequently, road users will quickly adapt to a new coordination and they will switch to faster routes if they suffer from last intervention. The traffic assignment, i.e. the distribution of traffic in a road network, may change significantly. In this paper, we study properties of this cyclically time-expanded model. Although this time-expanded model is given by a linear program, it can reproduce very realistic flow-dependent travel times for urban traffic networks with signal control. In particular, both typical convex link performance functions and the time-dependent behavior of traffic, necessary for traffic signal coordination, are realized in this model. Additionally, the propagation of platoons of cars that is essential for the minimization of waiting times in signal controlled road network is modeled.

**Related Work.** Our approach is different to other mathematical models for signal optimization, e.g. those proposed by Gartner et al. (1975), Serafini & Ukovich (1989) or Wünsch et al. (2008). In contrast, with our new model, we simultaneously determine the implied new traffic assignment in the network. Our approach is also different to heuristic methods like TRANSYT (Robertson, 1969), since our model allows applying strict mathematical optimization techniques. Traffic assignment on its own is also a well studied problem. There are strict mathematical approaches using network flow theory. This branch of research started with the seminal work of Ford and Fulkerson (1956) and recent result especially focus on various aspects of traffic, e.g. a new concept of fairness in networks with congestion (Jahn, Möhring, Schulz, & Stier Moses, 2005) and flows over time with load dependent transit times (Köhler & Skutella, 2005). More application-oriented concepts include dynamic traffic assignment (DTA) techniques, see e.g. Szeto & Lo (2006) and Chiu et al. (2010), and simulation based solutions (Nagel & Flötteröd, 2009).

In contrast, the combined problem of traffic signal optimization and traffic assignment is rarely studied. Allsop and Charlesworth (1977) recognized the feedback between an optimized coordination and traffic assignment. They proposed an iterative approach where signal timings are optimized with TRANSYT. Afterwards, an equilibrium traffic assignment is computed. These steps are repeated until no change in the coordination occurs. Obviously, using heuristic methods iteratively, one may only hope to derive a local optimum. Even worse, this approach may lead to a decline of network performance as shown by Dickson (1981). Despite the interaction of signal coordination and traffic assignment, only few research was done in the integrated optimization. For example, Chiou (2005) presented a bilevel formulation based on the approach of Allsop and Charlesworth. Recent results were also made by Bell and Ceylan (2004) as well as Teklu et al. (2007) using genetic programming.

This short overview also reveals one of the main conflicts in traffic optimization. On the one hand, the practitioner favors very realistic and dynamic traffic models. However, these sophisticated models do not allow for strict mathematical optimization techniques for instances of relevant size. Heuristics, e.g. genetic algorithms or line search strategies in non-convex settings, have to be used. They are very sensitive to local optima and neither provide guarantees on the gap towards optimal solution or on the convergence ratio. Thus, one may only hope to improve the present solution, but the actual optimal solution remains unknown. On the other hand, strict mathematical optimization approaches like linear programming often require several assumptions on the problem formulations which lead to very simplified traffic models. However, these strategies often provide much more insight, e.g. via the dual problem formulation. That is out of the solution we can directly derive quantified suggestions for an improvement of the network performance. Therefore, one has to find a compromise between a qualified optimality result and the model's accuracy.

**Our contribution.** In this paper we aim at the crossover between static and dynamic models for traffic assignment. Our main goal is a combined optimization approach for simultaneous traffic assignment and traffic signal coordination with realistic load dependent travel times which still allows applying strict mathematical optimization techniques.
Firstly, we shortly present static and dynamic traffic models and introduce our cyclically time-expanded network model. In Section 3, we show that this new discretized and linearized model yields load dependent travel times which resemble main properties of common link performance functions. Moreover, our model is also capable of capturing platoons of cars of quickly varying density as well as changing phases of traffic signals. We verify the applicability of our model by simulation with state-of-the-art simulation tools, namely VISSIM and MATSim.

Nevertheless, our model is still linear. Thus, traffic assignments can be computed very efficiently. Furthermore, we present an extension of our cyclically time-expanded model in Section 4, which allows for the simultaneous optimization of traffic signal coordination.

2. A Cyclically Time-Expanded Model for Traffic Assignment

2.1. A different view on travel times in inner-city road networks

For traffic signals and their coordination, a time-dependent model, capable of describing the time-offset between consecutive intersections, is a vital ingredient. Hence, we have to consider traffic flows over time. A (traffic) flow over time can formally be defined as a function \( f \colon A \times [0,T) \rightarrow \mathbb{R}_+^n \) on a network \( G = (V,A,u) \), where \( V \) is the set of nodes (intersections), \( A \) is the set of links (roads), and \( u \colon A \rightarrow \mathbb{R}_+^n \) is the capacity in the network, and \([0,T)\) is the time interval. Flow particles entering \( e \in A \) at time \( t \) arrive at the head of \( e \) at time \( t + t_e(f(e,t)) \), where \( t_e \colon \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n \) is the travel time function or link performance function. The amount of flow that passes a link \( e \) can be calculated as \( \int_0^T f(e,t) \, dt \). Furthermore, flow conservation has to ensure that flow cannot leave a node before it arrives there. Additionally, one may assume that flow can be stored in a node for some time representing queues at intersections. In detail, for each \( \bar{t} \in [0,T) \) and each non-terminal \( v \in V \),

\[
\sum_{e \in \delta^-(v)} \int_0^{t_e(f(v,t))} f(e,t) \, dt \geq \sum_{e \in \delta^+(v)} \int_0^{t_e(f(v,t))} f(e,t) \, dt.
\]

Flows over time are widely studied in the mathematical literature; Skutella (2009) provides a good overview of recent results. Many dynamic flow problems with travel times independent of the flow can be solved using time-expanded networks. Already Ford and Fulkerson introduced time-expanded networks in their seminal work on network flow theory (Ford & Fulkerson, 1956). To create an expanded network out of a simple network, for every node, several copies of this node are added to the graph, one for each desired time step. These nodes are connected by arcs, where the various copies of the vertices are connected according to the travel times of the original arcs. Additionally, arcs connecting consecutive copies of the same node allow waiting. Yet, time-expanded networks are rather inefficient if the time horizon \( T \) is large, since the number of time steps determines the number of network copies that have to be provided.

Dynamic traffic assignment (DTA) is used to model flow dependent travel times and to compute user equilibria according to Wardrop’s principle. This approach allows the usage of solving techniques of Control Theory or Calculus of Variations. However, even for medium size networks it is very difficult to derive analytical solutions. Therefore, to be able to compute at least numerical solutions in an iterative approach, time is discretized and several other parameters like the number of routes are limited.

Link performance functions, as suggested by the Bureau of Public Roads (BPR), are a widely accepted approach to model load-dependent travel times. However, these static link performance functions do not provide an interface to traffic signal coordination. They just support an a priori estimation of the expected average delay. Furthermore, to the best of our knowledge, approaches like DTA are not used to model dynamic flow in realistic scenarios of relevant size on such a fine timely level as needed for traffic signals.

However, one should look into the causes for flow dependence of travel times in urban areas. In inner-city road networks, the interactions between road users are of minor importance for the travel times. In comparison, traffic signals and the rather low speed limit have a much greater impact. Thus, for urban networks, we suggest a decomposition of travel time into pure transit time and waiting time at the intersection. In the following, the
transit time is assumed to be constant and the flow-dependent component of travel time is caused only by the delay at traffic signals.

2.2. The cyclically time-expanded network

With constant transit times, time-expanded networks come into play again. Even better, since pre-timed traffic signals have a periodic behavior, it is not necessary to use a full time horizon expansion. Instead, we suggest a cyclic time-expansion where we expand only a time interval of size of the cycle time $T$ of the traffic signals and use only $k \in \mathbb{N}$ time steps of size $t = \frac{T}{k}$.

**Definition 1 (Cyclically time-expanded network).** Let $G = (V, A, u)$ be a network with capacities $u: A \to \mathbb{N}$ and non-negative integral transit times $t_e$. For a given cycle time $\Gamma \in \mathbb{N}$, the corresponding cyclically time-expanded network $G^\Gamma = (V^\Gamma, A^\Gamma, u^\Gamma)$ is constructed as follows.

- For each node $v \in V$, we create $\Gamma$ copies $v_0, v_1, ..., v_{\Gamma-1}$, thus $V^\Gamma = \{v_t | v \in V, t \in \{0, ..., \Gamma - 1\}\}$.
- For each link $e = (v, w) \in A$, we create $\Gamma$ copies $e_0, e_1, ..., e_{\Gamma-1}$ where arc $e_t$ connects node $v_t$ to node $w_{(t+t_e) \mod \Gamma}$. Arc $e_t$ has capacity $u(e_t) = u(e)$.
- Additionally, there are waiting arcs from $v_t$ to $v_{t+1}$ $\forall v \in V$ and $\forall t \in \{0, 1, ..., \Gamma - 2\}$ and from $v_{\Gamma-1}$ to $v_0$.

Please note that there is almost no difference between transit links and waiting arcs in the model. On waiting arcs one only moves in time whereas on regular links one moves in space and time. On all arcs the cost is the travel time on this arc. Thus, waiting time is travel time on waiting arcs. Hence, both kinds of arcs are treated in the same way and we do not need to explicitly distinguish between them in the modeling of traffic assignment.

Now, traffic signals with their green and red phases can be modeled by setting the capacity of a link starting during the red phase to zero, i.e. $u(e_t) = 0$ for some $t$. A simple example of a cyclically time-expanded network consisting of two links and one traffic signal is shown in Figure 1. Additionally, capacities of the waiting arcs correspond to the maximum queue length on a link and they can be chosen accordingly.

3. Inherent Link Performance in the Cyclically Time-Expanded Network

So far, we have introduced a model with constant transit times on links that are independent of the flow on the link. From a practitioner’s point of view, this might seem too restrictive for a in the first moment. However, there is in fact a time-dependent behavior and hidden in the time expansion. We will now study this time dependency by an analysis of our model for a single road with a traffic signal. Furthermore, we support the observations by a simulation of traffic flow on this link.

Fig. 1. Cyclic time-expansion of a small network with two links and one traffic signal. The thin, gray arcs starting at the red phase of the signal are switched off and, thus, they have zero capacity.
In the cyclically time-expanded model, we can look at traffic flow in two ways. Firstly, we can examine each of the copies of an original link and the adjacent waiting arcs on their own. This provides a dynamic and detailed view on the traffic flow and we can observe flow particles moving through the network in a timely fashion. Secondly, we can virtually contract this flow over time back to the original link, i.e. we regard all copies of the links and the corresponding waiting arcs as a single link and compute the average travel time over the whole time horizon. In that way, we have a more static view on traffic flow on this contracted link and temporal details are lost, but we can compare the travel time to other static models, now.

In the following, we will use the first point of view to compute the average travel times for the second perspective. We will see that travel time on a contracted link depends on numerous parameters, e.g. the distribution of incoming flow values over time. Furthermore, flow particles on the same (contracted) link at different times will experience different travel times which is very obvious in the time-expanded model. We use link performance to denote the average travel time on a contracted link. We will not present a closed function for computing travel times. Nevertheless, we use the term inherent link performance function according to related static traffic assignment models with flow dependent travel times. In the following, we fix most of the parameters and restrict the inherent link performance function to a small subset of its domain to at least demonstrate the most important properties of this implicit link performance in the cyclically time-expanded model.

3.1. Development of flow-dependent waiting times

We consider a very small network which consists of a single road with a traffic signal in the middle similar to the network in Figure 1. Assume a set of flow values assigned to the cyclically time-expanded version of this scenario and a fixed signal coordination to be given. The total waiting time in this small network can be determined by multiplying flow value and transit time and summing over all waiting arcs. Increasing flow will increase the waiting time, because more flow will be assigned to the waiting arcs. Furthermore, due to the bounded capacities, the flow units on the waiting arcs may not leave completely at the first green outgoing link if the link does not provide enough capacity. Instead, the flow will have to use more waiting arcs until the accumulated flow is drained off. This relation is illustrated in Figure 2. Therefore, if the incoming flow is raised linearly on all copies of an arc, then the growth of the waiting time will not be linear but rather quadratic. More precisely, the obtained function is piecewise linear, but converges to a quadratic function if the length of the time steps in the expansion converges to zero.

In Figure 3(a), we present the relation between flow and average waiting time on a single expanded link in the cyclically time-expanded network, i.e. our network consists of only one road with a traffic signal at its end. The
traffic is uniformly distributed on all copies of this link, a traffic signal with a cycle time of 60 seconds and a red
time of 20 seconds is put at the end of the road and a free speed travel time of 10 seconds is assumed. Additionally, a capacity reduction from two lanes to one lane at the traffic signal was used. We now compute the average travel time in this scenario with respect to the flow value. The obtained travel time in Figure 3(a) demonstrates the capability of our waiting arc model: although using only constant travel times the inherent, implicit link performance functions of the model are not linear. Even better, they seem to resemble common standard link performance functions. Figure 3(a) visualizes the average travel time of a flow particle on the 'contracted' link. The individual travel time of a single flow unit depends on its arrival time at the traffic signal. It ranges from 10 seconds for flow units arriving at green with no waiting queue at the signal up to 30 seconds for road users arriving at the beginning of the red phase.

This analysis suggests that our model can capture flow dependent travel times similar to standard link performance functions. However, we assumed traffic to be uniformly distributed over time. This is rather unrealistic for inner-city traffic. In particular, traffic signals create platoons of cars. Thus, we extend our analysis to flows with quickly changing traffic density.

3.2. Travel times for platoons of cars

In the cyclically time-expanded network, platoons of cars can be modeled by different flow values on the copies of each particular link. Some copies may be used at full capacity, other links may not be used at all. Varying flow values can be interpreted as platoons of different lengths and densities. Thus, our model is capable of creating, splitting, merging, compressing, and stretching these platoons.

Let us return to the single road from above. Even in this small example, a platoon can change the characteristics of the average waiting time dramatically. A platoon of cars can be described by its length and its density. For simplicity, we fix the density to be constant, thus, the platoon length can be considered to be proportional to the average link flow. To visualize the occurring effects even better, we change the green time to 30 seconds and the capacity of the outgoing road is set to twice the capacity of the incoming road. The average travel time now depends on two parameters: the length of the platoon and the arrival time of the head of the platoon at the traffic signal.

In Figure 3(b) and 3(c), the calculated average travel time is plotted versus the platoon length. In 3(b), the first flow unit arrives at the intersection 10 seconds before the signal turns green. Due to the higher capacity of the outgoing road the traffic signal can be used to densify the platoon, e.g., the cars may leave on parallel lanes. The first car of the platoon has to wait for the longest time, the last cars may pass the signal without stopping. In (c), the first flow particle arrives 20 seconds after the signal turned green. Thus, small platoons can pass without stopping, long platoons are split by the traffic signal which turns red 10 seconds after the head of the platoon has passed.

![Fig. 3. (a) Computed average travel time with respect to flow on a single link in the cyclically time-expanded network. Incoming traffic is uniformly distributed over time; (b)&(c) Inherent link performance function for a platoon of cars in the cyclically time-expanded model. In (b), the platoon is arriving 10 seconds before the signal turns green. Due to the higher capacity of the outgoing road the traffic signal can be used to densify the platoon, e.g., the cars may leave on parallel lanes. The first car of the platoon has to wait for the longest time, the last cars may pass the signal without stopping. In (c), the first flow particle arrives 20 seconds after the signal turned green. Thus, small platoons can pass without stopping, long platoons are split by the traffic signal which turns red 10 seconds after the head of the platoon has passed.](image-url)
Fig. 4. Calculated average travel time for a road user in a platoon with respect to platoon length and arrival time at the traffic signal. The signal turns green at t=0 and red at t=30. The inherent link performance functions of Figure 4 are obtained as profile for t=50 and t=20. Note that for fixed platoon length one obtains waiting time functions very similar to those used by Wünsch (2008).

Obviously, the implicit travel times in our model are quite different from standard static link performance functions found in traffic literature (e.g. Sheffi (1984), Gartner et al. (2005), and Ortúzar & Willumsen (2006)). There, link performance functions are assumed to be convex and monotonically increasing. This assumption simplifies the analysis of user equilibria considerably. In contrast, the inherent link performance functions in the cyclic time-expanded model may be decreasing or concave on some intervals. However, this is no inaccuracy or disadvantage of our model. As we will show below, these characteristics of travel time functions can in fact be reproduced by state-of-the-art traffic simulation tools which supports our model.

To evaluate the travel times computed by the cyclically time-expanded model, we use two well established simulation tools, namely VISSIM and MATSim. VISSIM, established by ptv AG, is a state-of-the-art micro-simulation, featuring nearly every aspect of urban traffic. MATSim, mainly developed by TU Berlin and ETH Zurich, is an agent-based simulation. MATSim uses a simpler queuing model, but it supports the computation of equilibrium assignments for thousands of traffic participants. Figure 4 presents the described scenario and the corresponding measured travel times. As one can clearly see, the predicted travel times in Figure 3 fit remarkably well to the simulated travel times even without a careful calibration of our model.

Hence, one can conclude that the common assumptions on static link performance functions may be realistic in rural areas or on highways. But they are not accurate enough for signalized inner-city traffic. Moreover, it can be concluded that the cyclic time expansion is able to model both classical link performance and link performance for traffic signal coordination and platoons. However, we cannot present a closed formulation here, since this depends on various parameters, e.g. number, length and density of platoons, traffic signal settings, and capacity changes of the link. If most parameters fixed, the average travel time of a flow unit of a single platoon in

Fig. 5. Link performance for the same scenario as in Figure 3, but now simulated and measured with VISSIM. Note that only the average travel time is displayed. The actual time depends on the position of the car within the platoon.
our scenario is depicted in Figure 5 with respect to both parameters platoon length (i.e. traffic flow) and arrival time at the intersection (i.e. offset of the signal).

4. Traffic Signal Optimization in the Cyclically Time-Expanded Model

Up to now, we analyzed important properties of the cyclically time-expanded model. In the following, we shortly show how our model can be used for simultaneous traffic signal optimization. For a more detailed description of our approach, see Köhler & Strehler (2010). The cyclically time-expanded model is a fully linear model with constant travel times on the expanded links. Thus, we can use a standard network flow algorithm to efficiently compute a system optimal dynamic assignment, i.e. fast and exact combinatorial and mathematical programming tools can be applied. A very fine time discretization of time steps less than one second is easily achievable and quickly solvable even for large scenarios.

Theorem 1. (Köhler & Strehler, 2010) Using the cyclically time-expanded network the traffic assignment problem for a fixed traffic signal coordination and for a fixed time granularity can be solved efficiently.

Proof. The assignment problem can be formulated as a linear program (Ahuja, Magnanti, & Orlin, 1993), because all travel times in the expanded network are constant. Since the LP has polynomial size with respect to the input, i.e. the cyclically time-expanded network, it can be solved in polynomial time, e.g. by using the ellipsoid method.

In Section 2.2, we modeled traffic signals by setting capacities of outgoing links to zero. Instead of this fixed constraint, we can produce switchable traffic signals by multiplying the capacity with a binary decision variable. Furthermore, we can couple these decision variables as in logic-based signal controls (Friedrich, 2002). Adding these binary variables and constraints to the linear program leads to the following mixed integer program which allows simultaneous optimization of traffic signals and traffic assignment:

\[
\begin{align*}
\min & \sum_{e \in A} \sum_{\theta \in A} t_{e\theta} f_{\theta}(e) \\
\text{s.t.} &\quad 0 \leq \sum_{\theta \in A} f_{\theta}(e) = f(e) \leq u(e) \quad \forall e \in A \setminus E \\
&\quad \sum_{e \in \delta^{-}(v)} f_{\theta}(e) = \sum_{e \in \delta^{+}(v)} f_{\theta}(e) \quad \forall \theta \in \emptyset \quad \forall v \in V \\
&\quad f_{\theta}(s_{t}, s_{p}) = d_{\theta} \quad \forall \theta \in \Theta \\
&\quad \sum_{t=1}^{k} b_{i}^{n} = 1 \quad \forall n \in \{1, \ldots, N\} \\
&\quad f(e) \leq b^{n} Q^{*} u(e) \quad \forall e \in E \\
&\quad f(e) \geq 0, \ b^{n} \in \{0,1\}^{k}
\end{align*}
\]
Here, we minimize the overall travel time of all road users (1) with simultaneous offset optimization at $N$ intersections. We suggest not to use the number of stoppages or the total delay as a measure of performance, because in this case flow units may take long detours just to avoid stopping or waiting. That is, a performance index—as used by TRANSYT—is not suitable for simultaneous optimization of traffic assignment. Furthermore, we use arbitrary many commodities $\theta \in \Theta$ with origin $s_\theta$, destination $z_\theta$, and demand $d_\theta$. We also divide the link sets into normal roads and interior lanes at intersections $E \subseteq A$. Constraint (2) ensures the capacity bounds on the normal links, and equation (3) guarantees flow conservation for each commodity in each node. Equation (4) forces that the whole demand is sent through the network. Constraints (5) and (6) model the signals. With (5), we choose exactly one offset for each signalized intersection, i.e. exactly one binary variable is set to 1. The matrices $Q^e$ in (6) contain the logical constraints for safely crossing the intersection and the right side of (6) accordingly adjusts the capacities of the interior links.

We applied this approach to compute optimized signal settings in various real-world scenarios. For example, we considered the inner-city of Cottbus, a German town with about 100,000 inhabitants, with 32 signalized intersections. The scenario is presented in Figure 6. We started with an already optimized signal setting for a fixed assignment, i.e. the signals provide minimum delay for this fixed assignment. This assignment was computed by VISUM (ptv AG) and we applied our mixed integer program without rerouting to optimize the signal settings. Afterwards, our approach could reduce waiting time in the network by another 11% by rerouting traffic flow and resetting traffic signals. An interesting property of our optimized solutions is the creation of circulations around the city center. This reduces conflicts between opposing traffic streams and allows progressive signal settings for many road users.

5. Conclusion

In this paper, we focused on travel times in our cyclically time-expanded network. As a main insight, we have shown that even a completely linear model is capable of creating complex travel time functions. Simulation results show that the cyclically time-expanded model is realistic and can capture much more effects of inner-city traffic than static link performance functions can do. Even better, the cyclically time-expanded model can be used to optimize traffic assignment and traffic coordination simultaneously with exact mathematical programming techniques.
Acknowledgements

The authors thank Gregor Wünsch and Christian Liebchen who were involved in the first ideas of the cyclically time-expanded model. We also thank Kai Nagel (TU Berlin) and Klaus Nökel (ptv AG) for providing various insights into traffic simulation concepts and the corresponding simulation tools. This work was funded by the German Federal Ministry of Education and Research (BMBF).

References


