Flows over time with load-dependent transit times*

EKKEHARD KÖHLER** MARTIN SKUTELLA**

March 26, 2002

Abstract

More than forty years ago, Ford and Fulkerson studied maximum 'dynamic' s-t-flows in networks with fixed transit times on the arcs and a fixed time horizon. They showed that there always exists an optimal solution which sends flow on certain s-t-paths at a constant rate as long as there is enough time left for the flow along a path to arrive at the sink.

Although this result does not hold for the more general setting of flows over time with load-dependent transit times on the arcs, we prove that there always exists a provably good solution of this structure. Moreover, such a solution can be determined very efficiently by only one minimum convex cost flow computation on the underlying 'static' network. Finally, we show that the time-dependent flow problem under consideration is strongly NP-hard and even cannot be approximated with arbitrary precision in polynomial time, unless P=NP.

1 Introduction

Flow variation over time is an important feature in network flow problems arising in various applications such as road or air traffic control, production systems, communication networks (e. g., the Internet), and financial flows. Their common characteristic are 'dynamic' networks with capacities and transit times on the arcs. In contrast to static flow problems, flow values on arcs may change with time in these networks.

Another crucial phenomenon in many of those applications is that the time taken to traverse an arc varies with the current amount of flow on this arc. Since it is already a highly nontrivial problem to map these two aspects into an appropriate and tractable mathematical network flow model, there are hardly any algorithmic techniques known which are capable of providing reasonable solutions even for networks of rather modest size. The main aim of this paper is to provide new insights and algorithmic results which will hopefully turn out to have the potential to contribute to practically efficient solution methods.

Problem definition and notation. We consider a directed network $G = (V, E)$ with node set V and arc set E. There is a source node $s \in V$, a sink $t \in V$, and a positive demand value D. Our aim is to find a *quickest flow over time* which satisfies demand D within minimal time horizon (or makespan) T while respecting the following restrictions. Each arc $e \in E$ has a positive capacity u_e which is interpreted as an upper bound on the rate of flow entering e, i. e., a capacity per unit time. Moreover, arc e has an associated positive *transit time* τ_e.

*This is a revised version of the conference proceedings publication [12].

**Technische Universität Berlin, Fakultät II — Mathematik und Naturwissenschaften, Institut für Mathematik, Sekr. MA 6–1, Straße des 17. Juni 136, D – 10623 Berlin, Germany, Email: {ekoehler,skutella}@math.tu–berlin.de

Many real-world applications have the difficult but crucial property that the amount of time needed to traverse an arc of the network increases as the arc becomes more congested. Thus, we consider the situation where τ_e is not fixed but depends on the amount of flow currently sent through arc e.

In the setting of flows over time (also called *time-dependent flows* in the following), the flow conservation constraints require that, for any point in time θ and for any node $v \in V \setminus \{s\}$, the total inflow into node v until time θ is an upper bound on the total outflow out of node v until time θ. In particular, the fact that the inflow may exceed the outflow means that flow can be stored in nodes. However, for a flow over time with finite time horizon T we require that, for any node $v \in V \setminus \{s,t\}$, the total inflow into node v until time T is equal to the total outflow out of node v until time T.

Modeling flow-dependent transit times. The critical parameter for modeling temporal dynamics of time-dependent flows is the presumed dependency of the actual transit time τ_e on the current (and maybe also past) flow situation on arc e. Unfortunately, there is a tradeoff between the need of modeling this usually highly complex correlation as realistically as possible and the requirement of retaining tractability of the resulting mathematical program.

Due to the latter condition, many models in the literature rely on relatively simple functions. For example, the transit time of an arc is often treated as a function of only the flow rate at time of entry to the arc. However, in many cases this assumption is unrealistic since it does, for example, not preserve the first-in-first-out property encountered in most applications.

In contrast, a fully realistic model of flow-dependent transit times on arcs must take density, speed, and flow rate evolving along the arc into consideration [9]. Unfortunately, even the solution of mathematical programs relying on simplifying assumptions is in general still impracticable, i. e., beyond the means of state-of-the-art computers, for problem instances of realistic size (as those occuring in real-world applications such as road traffic control).

Existing models and results. In the following we discuss some approaches which can be found in the literature. For a more detailed account and further references we refer to [2, 13, 16, 17].

Merchant and Nemhauser [14] formulate a non-linear and non-convex program where time is being discretized. In their model, the outflow out of an arc in each time period solely depends on the amount of flow on that arc at the beginning of the time period. However, the non-convexity of their model causes analytical and computational problems. In [15] and [3] special constraint qualifications are described which are necessary to guarantee optimality of a solution in this model. Carey [4] introduces a slight revision of the model of Merchant and Nemhauser which transforms the non-convex problem into a convex one.

Carey and Subrahmanian [5] introduce a time-expanded network[1] with fixed transit times on the arcs. However, for each time period, there are several copies of an arc of the underlying 'static' network corresponding to different transit times. In this setting, flow-dependent transit times can implicitly be modeled by introducing appropriate capacities on the copies of an arc corresponding to different transit times. While the algorithmic techniques which can then be applied to the resulting time-expanded network are typically

[1]Time-expanded networks are often used for computing time-dependent flows with fixed (integral) transit times on the arcs. Such a time-expanded network contains a copy of the node set of the underlying 'static' network for each discrete time step (building a *time layer*). Moreover, for each arc with transit time τ in the static network, there is a copy between each pair of time layers of distance τ in the time-expanded network.

very efficient, the size of the time-expanded graph itself causes problems when the number of discrete time steps gets large.

This problem was already addressed by Ford and Fulkerson [7, 8] when they studied the 'maximal dynamic flow problem': Given a directed network with capacities and fixed transit times on the arcs and a fixed time horizon T, send as much flow as possible from the source vertex to the sink vertex within time T. This problem can be solved by one max-flow computation on the corresponding time-expanded network. Notice, however, that the size of the time-expanded network is only pseudo-polynomial in the size of the input.

Nevertheless, Ford and Fulkerson were able to show that the problem can be solved by essentially one min-cost flow computation on the underlying 'static' network, where transit times of arcs are interpreted as cost coefficients. An optimal solution to this min-cost flow problem can be turned into a flow over time by first decomposing it into flows on paths. The optimal time-dependent flow starts to send flow on each path at time zero, and repeats each so long as there is enough time left in the T periods for the flow along the path to arrive at the sink. Such a flow over time is called *temporally repeated*.

In particular, this result of Ford and Fulkerson implies that the time-dependent quickest flow problem with fixed transit times can be solved in polynomial time. However, the min-cost version of the quickest flow problem is already NP-hard [11]. On the other hand, Hoppe and Tardos [10] show that there is a non-trivial generalization of the result of Ford and Fulkerson to the case of multiple sources and sinks.

Recently, and prior to this work, Fleischer and Skutella [6] showed that the technique of Ford and Fulkerson can be generalized to yield approximate solutions to the NP-hard quickest flow problem with costs and also to the more general quickest multicommodity flow problem with costs. Their approach is based on the computation of a static *length-bounded flow*[2] which is then turned into a flow over time, similarly as in Ford and Fulkerson's result. This leads to an approximation algorithm with performance guarantee $2 + \varepsilon$ for the quickest multicommodity flow problem with costs. Moreover, Fleischer and Skutella introduce the concept of so-called *condensed* time-expanded networks. They are obtained by rounding transit times of arcs and scaling time such that the size of the resulting time-expanded network is polynomially bounded in the input size. As a consequence, one obtains fully polynomial-time approximation schemes for the quickest multicommodity flow problem with costs and for related problems.

It is interesting to note that, in a different but related context, Roughgarden and Tardos [18] recently presented surprising results on the quality of the so-called *user equilibrium*[3] compared to an optimal solution for a static traffic flow problem with load-dependent transit times.

New results and models. Our work is inspired by the results of Ford and Fulkerson [7, 8] and Fleischer and Skutella [6]. Although their techniques cannot directly be applied to the more general setting of flow-dependent transit times, we show that a similar approach yields provably good temporally repeated flows for the time-dependent quickest flow problem with load-dependent transit times.

This result is based on the following fairly general model of flow-dependent transit times. We assume that, at each point in time, the uniform speed on an arc depends only on the amount of flow or *load* which is currently on that arc. This assumption captures for

[2] A static (multicommodity) flow is called *length-bounded* if it can be decomposed into flows on paths whose transit times are bounded from above by a given value.

[3] The 'user equilibrium' is the state that is reached when every unit of flow selfishly chooses the minimum latency path from its source to its destination, given the arc congestion caused by the rest of the flow in the network. This situation typically occurs in traffic or communication networks which are not regulated by some central authority.

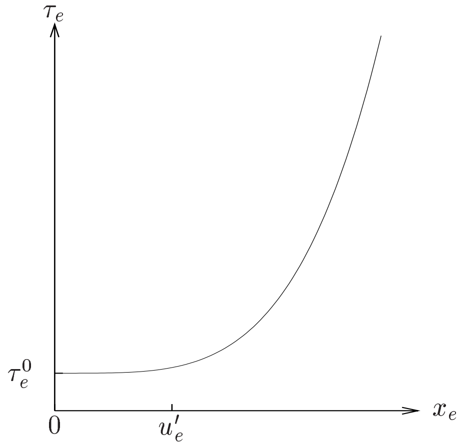

Figure 1: For the case of static road traffic, the U.S. Bureau of Public Roads developed a simplified function describing the dependency of the transit time on the flow. This function is given by $\tau_e = \tau_e^0(1 + 0.15(x_e/u'_e)^4)$, where τ_e^0 is the free-flow transit time, x_e is the flow rate, and u'_e is the 'practical capacity' of arc e. It follows from the given equation that the practical capacity of an arc is the flow rate at which the transit time is 15% higher than the free-flow transit time. More details can be found in [19].

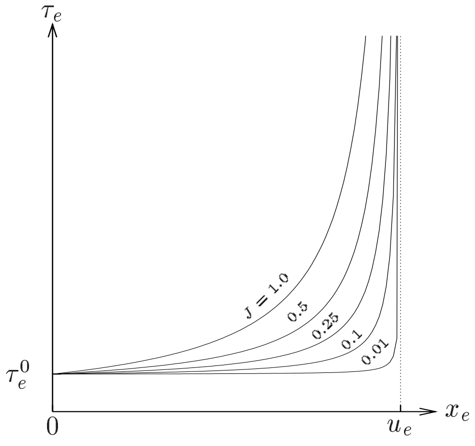

Figure 2: Davidson proposed a function describing the transit time based on queuing theory considerations: $\tau_e = \tau_e^0(1 + Jx_e/(u_e - x_e))$. Again, τ_e^0 is the free-flow transit time (i. e., the transit time at zero flow), x_e is the flow rate, u_e is the capacity of the arc, and J is a parameter of the model. We depict the function for various choices of J. In contrast to the function given in Figure 1, Davidson's function is asymptotic to the capacity u_e of the arc. Again, more details can be found in [19].

example the behavior of road traffic when an arc corresponds to a rather short street (notice that longer streets can be replaced by a series of short streets).

In static, i. e., non-time-varying flows, the load of an arc is uniquely determined by its flow rate which is the number of flow units traversing the arc per time unit. Therefore, the transit time of an arc is a function of its flow rate in this case. We assume that this dependency is given by an increasing and convex function. This prerequisite is satisfied for almost all applications, see, e. g., [19] and Figures 1 and 2.

We propose an algorithm which is similar to the one of Ford and Fulkerson and thus also very efficient. However, since the transit times are no longer fixed, the linear min-cost flow problem considered by Ford and Fulkerson now turns into a *convex cost* flow problem. Under the assumptions on the transit time functions τ_e stated above, the resulting optimal static flow can be turned into a temporally repeated flow which needs at most twice as long as a quickest flow over time.

The analysis of this algorithm is similar to the one of the $(2 + \varepsilon)$-approximation algorithm in [6]. The common underlying idea is to prove that averaging a flow over time with time horizon T yields a feasible static flow which can then again be turned into a flow over time with time horizon $2T$. However, while Fleischer and Skutella use the fact that the average static flow is length-bounded by T, this is no longer true if we allow flow-dependent transit times on the arcs.

Finally, we show that the time-dependent quickest flow problem under consideration is strongly NP-hard and even cannot be approximated with arbitrary precision in polynomial time, unless P=NP. We give reductions from the NP-hard problems PARTITION and SATIS-

FIABILITY. Notice that in contrast to this result the time-dependent flow problem considered by Ford and Fulkerson is solvable in polynomial time and that the problems considered by Fleischer and Skutella [6] possess fully polynomial-time approximation schemes and are therefore not strongly NP-hard, unless P=NP.

The paper is organized as follows. In Section 2 we discuss the load-dependent transit time model which our results are based on. Our main result on the existence and efficient computation of provably good temporally repeated flows is presented in Section 3. Section 4 contains our results on the complexity of the time-dependent flow problem under consideration.

2 A model for load-dependent transit times

Our research is motivated by traffic routing problems. Here, the transit time $\tau_e(x_e)$ of an arc (street) e is usually given as a function of the flow rate x_e which is the number of flow units (cars etc.) traversing the arc per time unit.

Unfortunately, it is not clear how to measure the rate of flow at a specific moment in time for a flow varying over time. For example, even if the number of cars which currently enter a street is small compared to the capacity of the street, their transit times might nevertheless be huge due to traffic congestion caused by a large number of cars which have entered the street earlier.

Therefore, the transit time of an arc is typically given for the case of static flows, that is, flows which do not vary over time. We thus interpret $\tau_e(x_e)$ as the transit time on arc e for the static flow rate x_e. However, since we are interested in the general setting where the flow on an arc may vary over time, we have to find a model which enables us to at least approximately determine transit times for time-dependent flows.

The small example of a congested street discussed above suggests that the transit time on an arc depends on its current load, which is the amount of flow (number of cars) currently on that arc. If we let y_e denote the load of arc e, it is easy to see that, for a static flow, the following relation holds:

$$y_e \;=\; x_e\,\tau_e(x_e) \;\;. \tag{1}$$

Observation 2.1. *If the function τ_e is monotonically increasing and convex, then, in a static flow, the flow rate x_e is a strictly increasing and concave function of the load y_e.*

Proof. Since both τ_e and the identity are non-negative, monotonically increasing, and convex, it follows from (1) that the load is also a non-negative, strictly increasing, and convex function of the flow rate x_e. Thus, the inverse function exists and is strictly increasing and concave. □

It follows from Observation 2.1 that, for the case of static flows, the transit time τ_e can also be interpreted as a function of the load y_e; to avoid ambiguity, it is then denoted by $\hat{\tau}_e(y_e)$. Notice that

$$\tau_e(x_e) = \hat{\tau}_e(y_e) \tag{2}$$

if the flow rate x_e and the load y_e satisfy equation (1). The function $\hat{\tau}_e$ can now be used in order to model the setting of flows over time with load-dependent transit times.

A time-dependent flow on arc e with time horizon T can be described by its flow rate $f_e : (0, T] \to \mathbb{R}^+$. There are several possible ways of defining the flow rate f_e depending on where exactly it is measured. For example, the *inflow rate* is measured at the tail and the *outflow rate* is measured at the head of an arc; moreover, a flow rate can also be determined at any other position on an arc.

The considerations and results in the remainder of this paper do not depend on the precise definition but work for all possible types of flow rates. However, it is a non-trivial task to determine, for example, the outflow rate of an arc from a given time-dependent inflow rate. Moreover, and in contrast to the situation for static flows given in (1), the same is true for the problem of determining the load $\ell_e : (0,T] \to \mathbb{R}^+$ of arc e from its time-dependent flow rate f_e. The mutual dependence of the flow rate and the load is implicitly given as follows. At any point in time θ, the speed of the flow on arc e is proportional to the inverse of the 'current transit time' $\hat{\tau}_e(\ell_e(\theta))$.

We are interested in two basic characteristics of flows over time on an arc e. The total transit time is the total amount of time spent by all units of flow on that arc. If we think of cars driving along a street, it is the sum of the individual transit times of all cars. Formally, the total transit time is given by

$$\int_0^T \ell_e(\theta)\, d\theta \quad .$$

The total amount of flow shipped through arc e is the integral over the flow rate $f_e(\theta)$, $0 < \theta \leqslant T$. This value can also be written in terms of the load $\ell_e(\theta)$:

$$\int_0^T f_e(\theta)\, d\theta \quad = \quad \int_0^T \frac{\ell_e(\theta)}{\hat{\tau}_e(\ell_e(\theta))}\, d\theta \quad . \tag{3}$$

Notice, however, that this equation does not hold pointwise (with the integrals removed) since the load $\ell_e(\theta_0)$ at some point in time θ_0 does not only depend on the current flow rate $f_e(\theta_0)$ but also on the whole history $f_e(\theta)$, $0 < \theta \leqslant \theta_0$.

For an arbitrary flow over time on arc e with time horizon T and flow rate $f_e(\theta)$, for $\theta \in (0,T]$, we define a corresponding static flow with flow rate

$$x_e \;:=\; \frac{1}{T}\int_0^T f_e(\theta)\, d\theta \tag{4}$$

and load $y_e := x_e\, \tau_e(x_e)$ according to (1). We refer to this static flow as the *average rate flow* corresponding to the flow over time given by f_e (or ℓ_e).

Lemma 2.2. *The total transit time of a time-dependent flow with time horizon T, flow rate f_e and load ℓ_e is at least as big as the total transit time of the corresponding static average rate flow over T time units, that is,*

$$\int_0^T \ell_e(\theta)\, d\theta \quad \geqslant \quad T\, y_e \quad . \tag{5}$$

Proof. The left hand side of (5) motivates the consideration of another static flow with load

$$\tilde{y}_e \;:=\; \frac{1}{T}\int_0^T \ell_e(\theta)\, d\theta$$

and flow rate

$$\tilde{x}_e \;:=\; \frac{\tilde{y}_e}{\hat{\tau}_e(\tilde{y}_e)} \quad .$$

We refer to this static flow as the *average load flow* corresponding to the flow over time given by f_e (or ℓ_e). Thus, (5) can now be rewritten as $\tilde{y}_e \geqslant y_e$. Since the load of a static flow is a monotonically increasing function of its flow rate, it suffices to show that $\tilde{x}_e \geqslant x_e$. Using (3) and (4), this inequality can be rewritten as

$$\frac{\tilde{y}_e}{\hat{\tau}_e(\tilde{y}_e)} \quad \geqslant \quad \frac{1}{T}\int_0^T \frac{\ell_e(\theta)}{\hat{\tau}_e(\ell_e(\theta))}\, d\theta \ . \tag{6}$$

It follows from Observation 2.1 that the function $\xi \mapsto \xi/\hat{\tau}_e(\xi)$ is concave. Thus, inequality (6) is a result of Jensen's inequality. This concludes the proof. ☐

In other words, Lemma 2.2 says that, for an arbitrary flow over time, the corresponding average load flow is stronger than the corresponding average rate flow.

3 An approximation algorithm

In order to determine a flow over time which satisfies demand D in close to optimal time, we consider the following static maximum flow problem with bounded convex cost. In this problem, the cost of flow x_e on arc e is $x_e\,\tau_e(x_e)$ and the total cost must not exceed D. More formally, the problem can be written as follows:

$$\begin{aligned}
\max \quad & \sum_{e\in\delta^-(t)} x_e - \sum_{e\in\delta^+(t)} x_e \\
\text{s.t.} \quad & \sum_{e\in\delta^-(v)} x_e - \sum_{e\in\delta^+(v)} x_e \;=\; 0 && \forall\, v \in V\setminus\{s,t\} \\
& \sum_{e\in E} x_e\,\tau_e(x_e) \;\leqslant\; D \\
& 0 \;\leqslant\; x_e \;\leqslant\; u_e && \forall\, e\in E
\end{aligned}$$

Here, $\delta^+(v)$ and $\delta^-(v)$ denote the set of arcs leaving node v and entering node v, respectively.

Lemma 3.1. *If there is a flow over time with load-dependent transit times which sends D units of flow from s to t within time T, then there exists a static flow of value at least D/T for the static flow problem stated above.*

Proof. Consider the static average rate flow $x = (x_e)_{e\in E}$ as defined in (4) which corresponds to the given flow over time with flow rate $(f_e)_{e\in E}$ and load $(\ell_e)_{e\in E}$. By the assumption on the amount of flow sent, the value of the static flow x is D/T. Moreover, since the given flow over time obeys the capacity constraints, the same is true for the average rate flow x. It thus remains to show that the cost of x is bounded by D.

Since in the time-dependent flow every unit of flow needs at most time T to travel through the network from s to t, the total transit time is bounded by $D\,T$, that is,

$$\sum_{e\in E} \int_0^T \ell_e(\theta)\, d\theta \quad \leqslant \quad D\,T \ .$$

By Lemma 2.2 and (1), this concludes the proof. ☐

An optimal integral solution to the static constrained maximum flow problem stated above can be computed in polynomial time, for example by the capacity scaling algorithm of Ahuja and Orlin [1]. Moreover, by appropriately scaling the data, we can make sure that the value of an optimal integral solution is arbitrarily close to the fractional optimum. In this way, we can compute a static flow x of value at least $(1-\varepsilon)$ opt and cost at most D in polynomial time, where opt is the value of an optimal static flow.

Lemma 3.2. *If there is a flow over time with load-dependent transit times which sends D units of flow from s to t within time T, then a static flow x of value at least $(1-\varepsilon)D/T$ and cost at most D can be computed in polynomial time.*

Proof. The result follows from Lemma 3.1 and the discussion above. □

Although it is a static flow, x contains some structural information on how to construct a provably good flow over time. We decompose x into a sum of static path-flows on a set of s-t-paths $\mathcal{P}$. The flow value on path $P \in \mathcal{P}$ is denoted by x_P such that, for each arc $e \in E$,

$$x_e = \sum_{P \in \mathcal{P} \,:\, e \in P} x_P \ .$$

Notice that we can assume without loss of generality that no cycles are needed in the flow decomposition; otherwise, the solution x can be improved by decreasing flow on those cycles. Moreover, it is well-known that the number of paths in $\mathcal{P}$ can be bounded by the number of arcs $|E|$.

A temporally repeated flow over time with time horizon T' can be generated from the path-decomposition of x by starting each path-flow at time zero, and repeating each so long as there is enough time left in the T' periods for the flow along the path to arrive at the sink.

As soon as we have computed the static flow x, we assume that the transit time of every arc $e \in E$ in the network is fixed to $\tau_e(x_e)$. This assumption is justified if we can assure that the rate of flow into arc e is always bounded by x_e and thus its load is never above $x_e\,\tau_e(x_e)$. In this case, we can enforce the constant transit $\tau_e(x_e)$ by introducing waiting times at the head node v of arc e in order to compensate for a potentially smaller transit time on that arc. Thereby we emulate the fixed transit time $\tau_e(x_e)$ and, at the same time, make sure that the rate of flow into every arc $e' \in \delta^+(v)$ also stays below $x_{e'}$. The same technique is used in [6] in order to round up transit times of arcs.

Under these assumptions, the amount of flow that can be sent within time horizon T' over path $P \in \mathcal{P}$ of length $\tau_P := \sum_{e\in P} \tau_e(x_e)$ with $\tau_P \leqslant T'$ is $x_P\,(T' - \tau_P)$. Therefore, the total amount of flow which we can send on paths $\mathcal{P}$ within time horizon T' is

$$d(T') := \sum_{P \in \mathcal{P} \,:\, \tau_P \leqslant T'} x_P\,(T' - \tau_P) \ . \tag{7}$$

From this expression one can easily determine the minimal T' that is needed to satisfy demand D: Simply order the paths in the set $\mathcal{P}$ by non-decreasing lengths $\tau_{P_1} \leqslant \tau_{P_2} \leqslant \cdots \leqslant \tau_{P_k}$ and observe that the function d is affine linear and increasing within every interval $[\tau_{P_i}, \tau_{P_{i+1}}]$, $1 \leqslant i < k$.

Theorem 3.3. *If there is a flow over time with load-dependent transit times which sends D units of flow from s to t within time T, then there exists a temporally repeated flow satisfying demand D within time horizon at most $2T$. Moreover, for every $\varepsilon > 0$, one can compute a temporally repeated flow in polynomial time which satisfies demand D within time horizon at most $(2+\varepsilon)T$.*

Proof. By Lemma 3.2, we can compute a static flow x of value at least $(1-\varepsilon/3)D/T$ and cost at most D in polynomial time, for any $\varepsilon > 0$ (in the following we assume that $\varepsilon \leqslant 1$). Moreover, by Lemma 3.1, there even exists such a flow with $\varepsilon = 0$. We decompose x into path-flows as discussed above and get

$$\begin{aligned} d\big((2+\varepsilon)\,T\big) &= \sum_{P\in\mathcal{P}\,:\,\tau_P\leqslant(2+\varepsilon)\,T} x_P\,\big((2+\varepsilon)\,T-\tau_P\big) \\ &\geqslant \sum_{P\in\mathcal{P}} x_P\,\big((2+\varepsilon)\,T-\tau_P\big) \\ &= (2+\varepsilon)\,T\sum_{P\in\mathcal{P}} x_P \;-\; \sum_{e\in E} x_e\,\tau_e(x_e) \\ &\geqslant (2+\varepsilon)\,(1-\varepsilon/3)\,D \;-\; D \\ &\geqslant D\ . \end{aligned}$$

Since the function d is increasing, this concludes the proof. □

It can be shown that our analysis is tight, even for the case of a network consisting of only one uncapacitated arc e from s to t. In this example, the transit time of arc e is given by

$$\tau_e(x_e) \;:=\; \begin{cases} 2 & \text{if } x_e \leqslant 1, \\ \infty & \text{if } x_e > 1. \end{cases}$$

From (2) one can determine the corresponding load-dependent function

$$\hat{\tau}(y_e) \;=\; \begin{cases} 2 & \text{if } y_e \leqslant 2, \\ \infty & \text{if } y_e > 2. \end{cases}$$

Thus, a quickest flow over time sending demand $D := 2$ from s to t needs exactly 2 time units since it can put the 2 units of flow onto arc e at time 0 such that they arrive at time 2. However, an optimal solution to the static maximum flow problem stated at the beginning of this section sets the flow rate x_e to 1. Thus, in the resulting temporally repeated flow the last piece of flow leaves s only at time 2 and therefore does not arrive before time 4.

In this example, the gap of 2 between the optimal solution and the solution arising from the static maximum flow problem obviously originates from the incapability of the static flow problem to capture the possibility of sending flow at a very high flow rate for a very short period of time.

Although the optimal flow over time is temporally repeated here, a slight modification of the instance shows that every temporally repeated flow can be bad compared to an optimal flow over time. If we double the demand value to $D := 4$, an optimal flow over time needs 4 time units since it sends two packets of flow, each containing 2 units of flow, at time 0 and at time 2, respectively. However, in a temporally repeated flow, the flow rate cannot be chosen bigger than 1 since otherwise the arc would become completely congested as soon as there are more than 2 units of flow on it. Therefore, every temporally repeated flow needs at least 6 time units and is thus at least a factor of $3/2$ away from the optimal value 4.

The results in Theorem 3.3 can be generalized in the following direction. One can decrease the factor of 2 in time at the cost of a decrease of the amount of flow that can be delivered. This leads to the following bicriteria results.

Corollary 3.4. *If there is a flow over time which sends D units of flow from s to t within time T, then, for every $\alpha \geqslant 1$, there exists a temporally repeated flow satisfying demand $(\alpha-1)D$ within time horizon at most αT. Moreover, for every $\varepsilon > 0$, one can* *compute a temporally repeated flow in polynomial time which satisfies demand* $(\alpha - 1)D$ *within time horizon at most* $(\alpha + \varepsilon)T$.

The proof of Corollary 3.4 is almost identical to the proof of Theorem 3.3 (just replace 2 by α).

We close this section with the following alternative view of the presented approximation result which highlights its close relation to the algorithm of Ford and Fulkerson. A temporally repeated flow with load-dependent transit times and time horizon T can be obtained from a solution to the following static convex cost flow problem.

$$\begin{aligned} \max \quad & T\left(\sum_{e\in\delta^-(t)} x_e - \sum_{e\in\delta^+(t)} x_e\right) - \sum_{e\in E} x_e\,\tau_e(x_e) \\ \text{s.t.} \quad & \sum_{e\in\delta^-(v)} x_e - \sum_{e\in\delta^+(v)} x_e \;=\; 0 \qquad \forall\, v \in V \setminus \{s,t\} \\ & 0 \;\leqslant\; x_e \;\leqslant\; u_e \qquad \forall\, e \in E \end{aligned}$$

For the special case of fixed transit times, this is exactly the static flow problem considered by Ford and Fulkerson. It is easy to observe that the value of the resulting temporally repeated flow is equal to the value of the objective function (compare (7)). For the special case of fixed transit times, Ford and Fulkerson showed that this temporally repeated flow over time is maximal.

This is no longer true in the setting with load-dependent transit times. However, it follows from Lemma 3.1 that for $T = 2T^*$ (where T^* denotes the makespan of a quickest flow) this value is at least D. Thus, we get an alternative $(2 + \varepsilon)$-approximation algorithm for the quickest flow problem by embedding this approach into a binary search framework for T. The main drawback of this alternative algorithm is that it requires more than one convex cost flow computation. Moreover, the simple example discussed above also shows that its performance guarantee is not better than 2.

4 Complexity results

While the corresponding problem with fixed transit times can be solved efficiently [7, 8], the quickest s-t-flow problem with load-dependent transit times is NP-hard. We start with a simple reduction from the well-known NP-complete PARTITION problem.

<u>PARTITION</u>

Given: A set of n items with associated sizes $a_1, \dots, a_n \in \mathbb{N}$ such that $\sum_{i=1}^n a_i = 2L$ for some $L \in \mathbb{N}$.

Question: Is there a subset $I \subset \{1, \dots, n\}$ with $\sum_{i\in I} a_i = L$?

Given an instance of PARTITION, we construct a network with load-dependent transit times as follows. Take a chain of length n where each link $i = 1, \dots, n$ consists of a pair of two parallel arcs e_i and $\bar{e}_i$ (see Figure 3) with the following transit times:

$$\tau_{e_i}(x) \;:=\; \begin{cases} 2L & \text{if } x \leqslant 1/(2L), \\ \infty & \text{if } x > 1/(2L), \end{cases}$$

$$\hat{\tau}_{e_i}(y) \;=\; \begin{cases} 2L & \text{if } y \leqslant 1, \\ \infty & \text{if } y > 1, \end{cases}$$

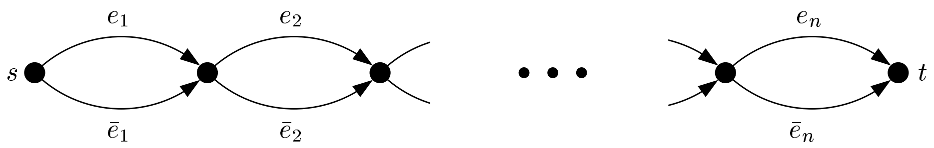

Figure 3: Reduction of the problem PARTITION to a time-dependent flow problem with load-dependent transit times.

and

$$\tau_{\bar{e}_i}(x) := \begin{cases} 2L + a_i & \text{if } x \leqslant 1/(2L + a_i), \\ \infty & \text{if } x > 1/(2L + a_i), \end{cases}$$

$$\hat{\tau}_{\bar{e}_i}(y) = \begin{cases} 2L + a_i & \text{if } y \leqslant 1, \\ \infty & \text{if } y > 1. \end{cases}$$

The task is to send $D := 2$ units of flow from s to t.

Lemma 4.1. *There exists a flow over time which sends two units of flow from s to t in time $(2n + 1)L$ if and only if the underlying instance of* PARTITION *is a 'yes'-instance.*

Proof. If: Let I be a subset of $\{1, \dots, n\}$ such that $\sum_{i \in I} a_i = L$. The flow is sent in two packets, each containing one flow unit. The packets use two arc-disjoint paths of length (transit time) $(2n + 1)L$ that are induced by the subset I and its complement, respectively.

Only if: It follows from the definition of the transit time functions that there can never be more than one unit of flow on any arc. Therefore the construction of the network yields that no arc can be traversed by more than one flow unit unless the flow takes at least $(2n + 2)L$ units of time. As a consequence, in a flow over time with makespan $(2n+1)L$, every arc is traversed by exactly one flow unit and the total transit time is thus $\sum_{i=1}^{n}(2L + 2L + a_i) = 2(2n + 1)L$. In particular, an arbitrary piece of flow needs exactly $(2n + 1)L$ units of time to travel from s to t and the corresponding path therefore induces a subset $I \subset \{1, \dots, n\}$ with $\sum_{i \in I} a_i = L$. □

So far we have shown that the problem under consideration is NP-hard in the weak sense. Next we give a more involved reduction from the NP-complete SATISFIABILITY problem.

SATISFIABILITY

Given: n Boolean variables $z_1, \dots, z_n$ and m disjunctive clauses $C_1, \dots, C_m$.

Question: Does there exist a truth-assignment which satisfies all clauses?

The aim of the following reduction is to create a gap between those instances of the flow problem corresponding to 'yes'-instances of SATISFIABILITY and those corresponding to 'no'-instances. This gap then yields a non-approximability result for the flow problem under consideration.

For every variable z_i of the SATISFIABILITY instance, we introduce two outgoing arcs e_i and $\bar{e}_i$ from the source s and one ingoing arc a_i to the sink t. Moreover, for every clause C_j, there is an ingoing arc c_j to the sink. There are additional arcs from the head of e_i and $\bar{e}_i$ to the tail of a_i. Finally, for every variable z_i occuring unnegated (negated) in clause C_j, there is an arc from the head of e_i ($\bar{e}_i$) to the tail of c_j.

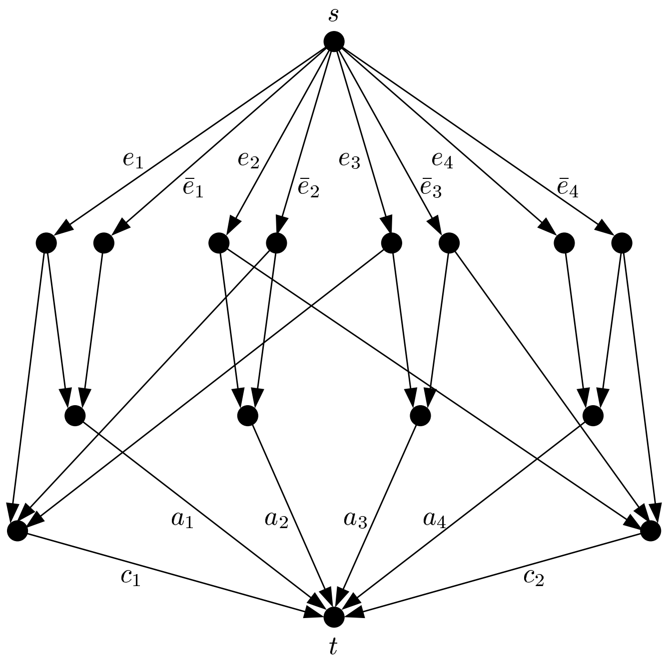

Figure 4: Reduction of the problem SATISFIABILITY to a time-dependent flow problem with load-dependent transit times. In the depicted example, the instance of SATISFIABILITY contains four variables z_1, z_2, z_3, and z_4 and two clauses $C_1 = (z_1 \vee \neg z_2 \vee z_3)$ and $C_2 = (z_2 \vee \neg z_3 \vee \neg z_4)$.

An illustration of this construction is given in Figure 4. All arcs are uncapacitated and have the following transit times (let $0 < \varepsilon \ll 1/(9m)$ be a small constant):

$$\begin{aligned}
\tau_{e_i}(x) &:= \tau_{\bar{e}_i}(x) := \begin{cases} \frac{1}{1-x} & \text{if } x < 1, \\ \infty & \text{if } x \geqslant 1, \end{cases} \\
\hat{\tau}_{e_i}(y) &= \hat{\tau}_{\bar{e}_i}(y) = 1 + y, \\
\tau_{a_i}(x) &:= \begin{cases} 2 & \text{if } x \leqslant 1/2, \\ \infty & \text{if } x > 1/2, \end{cases} \\
\hat{\tau}_{a_i}(y) &= \begin{cases} 2 & \text{if } y \leqslant 1, \\ \infty & \text{if } y > 1, \end{cases} \\
\tau_{c_j}(x) &:= \begin{cases} 3 & \text{if } x \leqslant \varepsilon/3, \\ \infty & \text{if } x > \varepsilon/3, \end{cases} \\
\hat{\tau}_{c_j}(y) &= \begin{cases} 3 & \text{if } y \leqslant \varepsilon, \\ \infty & \text{if } y > \varepsilon. \end{cases}
\end{aligned}$$

The transit times of all remaining arcs are fixed to 0.

Lemma 4.2. *If the underlying instance of* SATISFIABILITY *is a 'yes'-instance, then there* *exists a flow over time which sends $n+\varepsilon m$ units of flow from s to t in time $4+\varepsilon m$. However, if it is a 'no'-instance, then every flow over time needs at least $4 + 1/9$ units of time.*

Proof. Given a satisfying truth-assignment for the underlying instance of SATISFIABILITY, we construct a flow over time as follows. For every $i = 1, \dots, n$, if the variable z_i is set to true (false), then we route one unit of flow from s over $\bar{e}_i$ (e_i) and a_i to t. Since this is the only flow routed across these arcs, it will arrive at t at time 4.

For every clause C_j, $j = 1, \dots, m$, we choose a literal ℓ_j which is set to true. If ℓ_j is the unnegated (negated) variable z_i, then we route ε units of flow from s over e_i ($\bar{e}_i$) and c_j to t. Since at most εm units of flow are routed across e_i ($\bar{e}_i$) and exactly ε units of flow are routed across c_j, this flow arrives at t at time $4+\varepsilon m$. We have thus constructed a flow over time which sends $n + \varepsilon m$ units of flow from s to t in time $4 + \varepsilon m$.

On the other hand, we have to show that the existence of a flow over time with makespan less than $4 + 1/9$ yields a satisfying truth-assignment for the underlying instance of SATISFIABILITY. We first claim that, in such a flow, exactly one unit of flow is sent to t across arc a_i, for every $i = 1, \dots, n$, and exactly ε units of flow are sent over arc c_j, for every $j = 1, \dots, m$.

Notice that every unit of flow which is sent across a_i enters this arc between time 1 and $2 + 1/9$. Since the transit time of a_i is at least 2, these flow units can simultaneously be found on the arc at time $2 + 1/9$. The construction of $\hat{\tau}_{a_i}$ yields that the total amount of flow which is sent across a_i is bounded by 1. A similar argument shows that at most ε units of flow can be sent to t across arc c_j, for every $j = 1, \dots, m$. Since exactly $n + \varepsilon m$ units of flow are sent from s to t, these bounds are tight which proves the claim.

In the following, we refer to the unit of flow sent across arc a_i as *commodity* i. For every $i = 1, \dots, n$, if at most one half of commodity i is sent across e_i, we set variable z_i to true; otherwise, we set it to false. It remains to show that this is a satisfying truth-assignment.

Consider some clause C_j and the flow which uses the corresponding arc c_j; we refer to this flow as *commodity* j. Moreover, consider an arc e leaving the source s which is used by commodity j. By construction of the network, e corresponds to a literal (z_i or $\neg z_i$) of clause C_j. It therefore suffices to show that the amount of flow of commodity i which is sent across arc e is less than $1/2$. In this case, the corresponding literal in clause C_j is set to true such that the clause is fulfilled.

Notice that, in order to arrive in time at the sink t, all flow of commodity j using arc e must arrive at the head of e before time $10/9$. In particular, arc e must not be congested too much in the time interval $[0, 10/9)$, that is, the 'speed' $1/\hat{\tau}_e(\ell_e(\theta))$ has to be large enough such that flow of commodity j can arrive at the head of e before time $10/9$. More formally, this implies the following condition on the load $\ell_e(\theta)$ of arc e:

$$\int_0^{10/9} \frac{1}{\hat{\tau}_e(\ell_e(\theta))} \, d\theta \;\geqslant\; 1 \; . \tag{8}$$

By contradiction, we assume that the amount d of flow of commodity i which is sent across arc e is at least $1/2$. In order to arrive in time at the sink t, all flow of commodity i using arc e must arrive at the head of e before time $19/9$. This yields

$$\int_0^{19/9} \frac{\ell_{e,i}(\theta)}{\hat{\tau}_e(\ell_e(\theta))} \, d\theta \;\geqslant\; d \; , \tag{9}$$

where $\ell_{e,i}(\theta)$ denotes the load of commodity i on arc e at time θ (also compare (3)).

Since $\ell_{e,i}(\theta) \leqslant \min\{\ell_e(\theta), d\}$, for all θ, and $d \geqslant 1/2$, we get

$$\begin{aligned}
\int_{10/9}^{19/9} \frac{\ell_{e,i}(\theta)}{\hat{\tau}_e(\ell_e(\theta))}\, d\theta &\leqslant \int_{10/9}^{19/9} \frac{\ell_{e,i}(\theta)}{1+\ell_{e,i}(\theta)}\, d\theta \\
&\leqslant \int_{10/9}^{19/9} \frac{d}{1+d}\, d\theta \\
&= \frac{d}{1+d} \leqslant \frac{2d}{3} .
\end{aligned}$$

The second inequality is based on the fact that the function $\xi \mapsto \xi/(1+\xi)$ is monotonically increasing in the interval $[0, \infty)$.

Together with (9), this yields

$$\int_0^{10/9} \frac{\ell_{e,i}(\theta)}{\hat{\tau}_e(\ell_e(\theta))}\, d\theta \geqslant \frac{d}{3} . \tag{10}$$

Putting together (8) and (10), we get the following contradiction:

$$\begin{aligned}
\frac{10}{9} &= \int_0^{10/9} \frac{\ell_e(\theta)}{\hat{\tau}_e(\ell_e(\theta))}\, d\theta + \int_0^{10/9} \frac{1}{\hat{\tau}_e(\ell_e(\theta))}\, d\theta \\
&\geqslant \int_0^{10/9} \frac{\ell_{e,i}(\theta)}{\hat{\tau}_e(\ell_e(\theta))}\, d\theta + 1 \\
&\geqslant \frac{d}{3} + 1 \geqslant \frac{7}{6} .
\end{aligned}$$

This concludes the proof. □

As a consequence of Lemma 4.2 we get the following hardness result for the flow problem under consideration.

Theorem 4.3. *The problem of finding a quickest flow over time with load-dependent transit times is strongly NP-hard and also APX-hard, i. e., there does not exist a polynomial-time approximation scheme, unless P=NP.*

Lemma 4.2 even yields a stronger non-approximability result.

Corollary 4.4. *There does not exist an approximation algorithm with performance guarantee better than* $37/36$*, unless P=NP.*

Certainly, stronger bounds than the one stated in Corollary 4.4 can be obtained using the same reduction with a more careful choice of the crucial parameters. However, we did not pursue this idea further.

Notice that the load-dependent transit times used for the hardness results in this section and the negative results of the last section are artificial and probably not very realistic. However, similar results can be obtained for more natural transit time functions. Unfortunately, the analysis gets much more involved then.

5 Concluding remarks

We have presented a very efficient technique for obtaining approximate solutions to the quickest flow problem with load-dependent transit times. We hope that this will also lead to practically efficient and useful algorithms. In particular, we plan to implement our algorithm and to test it on real-world data (road traffic networks).

The following interesting generalization of the quickest flow problem under consideration was pointed out by Lisa Fleischer (personal communication, January 2002). There are k commodities $i = 1, \ldots, k$, each given by a source-sink pair (s_i, t_i). The aim is to find a quickest multicommodity flow over time such that the sum of the flow values of all commodities is at least D. The approximation result described in Section 3 can directly be generalized to this setting; the static flow problem stated at the beginning of Section 3 is turned into a maximum multicommodity flow problem with cost bounded by D. The analysis remains essentially unchanged.

A problem closely related to the quickest flow problem is the maximum time-dependent flow problem with fixed time horizon T. Unfortunately, Corollary 3.4 does not yield any useful result for this variant of the problem. On the other hand, the reduction from PARTITION in Section 4 shows, that the problem cannot be approximated with performance guarantee strictly better than $1/2$, unless P=NP.

Acknowledgments. The authors are grateful to Lisa Fleischer for numerous helpful comments on this manuscript. Moreover, the second author is much indebted to Lisa Fleischer for interesting and helpful discussions on the subject of flows over time in general. While the immediate outcome of this cooperation can be found in [6], it also partly inspired the work at hand. Finally, the authors want to thank Arie Tamir for pointing out the connection to Jensen's inequality in the proof of Lemma 2.2.

References

[1] R. K. Ahuja and J. B. Orlin. A capacity scaling algorithm for the constrained maximum flow problem. *Networks*, 25:89–98, 1995.

[2] J. E. Aronson. A survey of dynamic network flows. *Annals of Operations Research*, 20:1–66, 1989.

[3] M. Carey. A constraint qualification for a dynamic traffic assignment model. *Transportation Science*, 20:55–58, 1986.

[4] M. Carey. Optimal time-varying flows on congested networks. *Operations Research*, 35:58–69, 1987.

[5] M. Carey and E. Subrahmanian. An approach to modelling time-varying flows on congested networks. *Transportation Research B*, 34:157–183, 2000.

[6] L. Fleischer and M. Skutella. The quickest multicommodity flow problem. In *Proceedings of the 9th Conference on Integer Programming and Combinatorial Optimization*, Lecture Notes in Computer Science. Springer, Berlin, 2002. To appear.

[7] L. R. Ford and D. R. Fulkerson. Constructing maximal dynamic flows from static flows. *Operations Research*, 6:419–433, 1958.

[8] L. R. Ford and D. R. Fulkerson. *Flows in Networks*. Princeton University Press, Princeton, NJ, 1962.

[9] N. Gartner, C. J. Messer, and A. K. Rathi. Traffic flow theory: A state of the art report. http://www-cta.ornl.gov/cta/research/trb/tft.html, 1997.

[10] B. Hoppe and É. Tardos. The quickest transshipment problem. *Mathematics of Operations Research*, 25:36–62, 2000.

[11] B. Klinz and G. J. Woeginger. Minimum cost dynamic flows: The series-parallel case. In E. Balas and J. Clausen, editors, *Integer Programming and Combinatorial Optimization*, volume 920 of *Lecture Notes in Computer Science*, pages 329–343. Springer, Berlin, 1995.

[12] E. Köhler and M. Skutella. Flows over time with load-dependent transit times. In *Proceedings of the 13th Annual ACM–SIAM Symposium on Discrete Algorithms*, pages 174–183, San Francisco, CA, 2002.

[13] H. S. Mahmassani and S. Peeta. System optimal dynamic assignment for electronic route guidance in a congested traffic network. In N. H. Gartner and G. Improta, editors, *Urban Traffic Networks. Dynamic Flow Modelling and Control*, pages 3–37. Springer, Berlin, 1995.

[14] D. K. Merchant and G. L. Nemhauser. A model and an algorithm for the dynamic traffic assignment problems. *Transportation Science*, 12:183–199, 1978.

[15] D. K. Merchant and G. L. Nemhauser. Optimality conditions for a dynamic traffic assignment model. *Transportation Science*, 12:200–207, 1978.

[16] W. B. Powell, P. Jaillet, and A. Odoni. Stochastic and dynamic networks and routing. In M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, editors, *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*, chapter 3, pages 141–295. North–Holland, Amsterdam, The Netherlands, 1995.

[17] B. Ran and D. E. Boyce. *Modelling Dynamic Transportation Networks*. Springer, Berlin, 1996.

[18] T. Roughgarden and É. Tardos. How bad is selfish routing? In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 93–102, Redondo Beach, CA, 2000.

[19] Y. Sheffi. *Urban Transportation Networks*. Prentice-Hall, New Jersey, 1985.