On the Performance of User Equilibria in Traffic Networks

Andreas S. Schulz * Nicolás Stier Moses *

1 Introduction

While Wardrop (1952) had introduced the concept of Nash equilibrium to *describe* user behavior in traffic networks, traffic engineers have proposed to utilize user equilibria in route-guidance systems to *prescribe* user behavior. Yet, Nash equilibria in general and user equilibria in particular are known to be inefficient and critics favored in principle the difficult-to-implement system optimum, which guarantees that the average travel time is minimal. Hence, the recent result that user equilibria are near optimal (Roughgarden and Tardos 2002) came as a welcome surprise, which helped to justify the use of user equilibria in retrospect.

In this paper, we extend the work of Roughgarden and Tardos in two directions. First, we introduce and analyze user equilibria in capacitated networks. In contrast to uncapacitated networks (the framework of Roughgarden and Tardos' work), the set of user equilibria is no longer convex and an equilibrium can be arbitrarily worse than the system optimum, even if arc latency functions are linear. Yet, we show that adding capacities does not change the worst-possible ratio between the *best* user equilibrium and the system optimum, given a fixed but arbitrary class of allowable latency functions. Second, it is well known that the system optimum can assign some drivers to paths having a significantly higher latency compared to other paths between the same origin-destination pair. For that reason, system optima are often considered inadequate for purposes of traffic planning. We propose to compare user equilibria to a more restricted definition of a system optimum, which is fairer in a certain sense and which therefore promises improved general acceptance. We study the quality of user equilibria in this setting and analyze the resulting improvement of their performance.

2 The Model

We consider a directed network $G = (V, A)$, and a set of k origin-destination pairs (s_i, t_i) with $s_i, t_i \in V$, $i = 1, \ldots, k$. A flow of rate d_i must be routed from s_i to t_i. Let P_i be the set of directed paths from s_i to t_i in G. Every arc a has a non-negative and non-decreasing *latency function* $\ell_a(\cdot)$, which maps the flow on a to the time needed to traverse a. We denote an instance of such a system by (G, d, ℓ). A flow f is a function that assigns a non-negative value to every path $P \in \mathcal{P}$, where $\mathcal{P} \equiv \cup P_i$. Given a flow, we can easily compute the utilization of an arc by $f_a = \sum_{P:a\in P} f_P$. A flow is feasible when $\sum_{P\in P_i} f_P = d_i$. The travel time along a path P is $\ell_P(f) \equiv \sum_{a\in P} \ell_a(f_a)$. To compare different flows, we use the total travel time in the network as a measure of efficiency. With our notation, it can be represented as $C(f) \equiv \sum_{P\in\mathcal{P}} \ell_P(f) f_P = \sum_{a\in A} \ell_a(f_a) f_a$. A feasible flow f^* that minimizes $C(f)$ is called a *system optimum.* A feasible flow f is a *user equilibrium* if for all origin-destination pairs (s_i, t_i) and for all $P, Q \in P_i$ such that $f_P > 0:\ \ell_P(f) \le \ell_Q(f)$ (Wardrop 1952).

3 User Equilibria in Capacitated Networks

We extend the basic model to include arc capacities c_a. Consequently, we need to define "user equilibrium". The natural extension is to assume that no user can switch to a shorter route *with residual capacity.* The following definition captures this assumption on users' behavior.

Definition 3.1. *A flow f is a* capacitated user equilibrium *if for all i and $P, Q \in P_i$ such that $f_P > 0$ and* $\min_{a\in Q} \{c_a - f_a\} > 0 : \ell_P(f) \le \ell_Q(f)$.

Because Definition 3.1 does not impose any restriction on paths that are used up to capacity, flow-carrying paths between the same origin-destination pair can have differing latencies, in contrast to the uncapacitated case. An important effect of arc capacities is therefore the existence of multiple equilibria. (In the uncapacitated case, the equilibrium is essentially unique.) The loss of uniqueness can be explained by the saturation of arcs, which restricts the choice for the remaining travelers. Figure 1 provides a simple example.

3.1 Beckman User Equilibrium We would like to characterize the best user equilibrium, but this is difficult because the set of all user equilibria is a non-convex region in the space of flows. Instead, we pick a particular user equilibrium that has a good characterization. We

*M.I.T., 77 Massachusetts Avenue, Cambridge, MA 02139, USA. Email: {schulz,nstier}@mit.edu

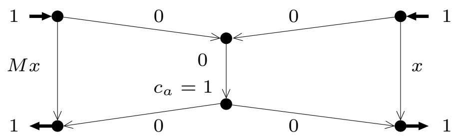

Figure 1: *Example with multiple equilibria.*

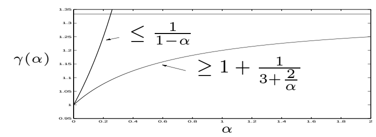

Figure 2: *Bounds for* $\gamma(\alpha)$.

define the *Beckman user equilibrium* (*BUE* for short) to be the solution to the following mathematical program.

$$\begin{array}{llcll} \min & \sum_{a\in A} \int_0^{f_a} \ell_a(x)dx & & & \\ \text{s.t.} & \sum_{P\ni a} f_P & = & f_a & \forall a \in A, \\ & \sum_{P\in P_i} f_P & = & d_i & \forall i = 1,\ldots,k, \\ & f_a & \leq & c_a & \forall a \in A, \\ & f_P & \geq & 0 & \forall P \in \mathcal{P}. \end{array}$$

As this problem is convex, its optimum can be described by first order conditions. Namely, a flow f is a *BUE* if and only if for all feasible circulations g:

$$\sum_a g_a \ell_a(f_a) \geq 0. \tag{3.1}$$

It is not difficult to show that this condition implies that every *BUE* is a user equilibrium in the sense of Definition 3.1.

3.2 Best and Worst Equilibria Roughgarden and Tardos (2002) observed that the worst-possible ratio between the total latency of the user equilibrium and the system-optimal latency can be arbitrarily large unless additional structure is imposed on the class of allowable arc latency functions or on the allowable network topology. Roughgarden (2002) showed that the network topology plays actually no role in the determination of this ratio. The example in Fig. 1 shows that capacities play a role; the travel time of the worst user equilibrium is not within a constant factor of the system optimum even though all arc latency functions are linear. However, we can use Eqn. (3.1) to adapt the proof of Roughgarden (2002) to show the following result.

THEOREM 3.1. *Let $\mathcal{L}$ be a standard class of latency functions. The worst-case ratio between the latency of the BUE and the latency of a system optimum is achieved by an uncapacitated network.*

In particular, the ratio between the best user equilibrium and the system optimum in a capacitated network with linear latency functions is at most 4/3.

4 Constrained System Optima

To overcome the intrinsic drawback of the system optimum, Jahn et al. (2000) proposed to consider the α-constrained system optimum, defined as an optimal solution to minimizing total travel time in a system where all users are assigned to paths of length not greater than $(1+\alpha)$ times their shortest paths. This concept allows for a more realistic assessment of the quality of user equilibria. Let f be an uncapacitated user equilibrium and let f^*_α be a constrained system optimum, for $\alpha \geq 0$. We introduce the function $\gamma(\alpha) \equiv \sup_{(G,d,\ell)} C(f)/C(f^*_\alpha)$. It is easy to verify for linear latency functions that $1 \leq \gamma(\alpha) \leq \frac{4}{3}$, for all α. Yet, these bounds can be improved.

THEOREM 4.1. *For linear latencies and $0 \leq \alpha < 1$, $1 + \frac{1}{3+\frac{2}{\alpha}} \leq \gamma(\alpha) \leq \frac{1}{1-\alpha}$. In particular, $\gamma(0) = 1$.*

These bounds are displayed in Fig. 2. In the limit, we recover the bounds of Roughgarden and Tardos. While the lower bound in Theorem 4.1 can be proved with the help of a collection of instances based on a modified Braess paradox network (Murchland 1970), we use polynomials of large degree for the general case.

LEMMA 4.1. *For arbitrary latency functions and $\alpha \geq 0$, $\gamma(\alpha) \geq \max\{1+\alpha, 2\}$.*

Finally, the following theorem implies that $\gamma(\alpha)$ is subadditive.

THEOREM 4.2. *The function $\gamma(\alpha)/\alpha$ is non-increasing.*

References

Jahn, O., R. Möhring, and A. Schulz (2000). Optimal routing of traffic flows with length restrictions in networks with congestion. In *Operations Research Proceedings 1999*, pp. 437–442. Springer.

Murchland, J. (1970). Braess's paradox of traffic flow. *Transportation Research 4*, 391–394.

Roughgarden, T. (2002). The Price of Anarchy is Independent of the Network Topology. *Manuscript*. To appear in Journal of Computer and System Sciences.

Roughgarden, T. and E. Tardos (2002). How bad is selfish routing? *Journal of the ACM 49*(2), 236–259.

Wardrop, J. (1952). Some theoretical aspects of road traffic research. In *Proceedings of the Institution of Civil Engineers*, Volume 1, pp. 325–378.