Equilibrium, Games, and Pricing in Transportation and Telecommunication Networks

EITAN ALTMAN
INRIA, 2004 Route des Lucioles B.P. 93, 06902 Sophia Antipolis, France
email: altman@sophia.inria.fr

LAURA WYNTER
IBM Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598, USA
email: lwynter@us.ibm.com

Abstract

Network equilibrium models that have traditionally been used for transportation planning have penetrated in recent years to other scientific fields. These models have recently been introduced in the telecommunication networks literature, as well as in the field of game theory. Researchers in the latter fields are not always aware of the very rich literature on equilibrium models outside of their application area. On the other hand, researchers that have used network equilibrium models in transportation may not be aware of new application areas of their tools. The aim of this paper is to present some central research issues and tools in network equilibria and pricing that could bring closer the three mentioned research communities.

Keywords: Wardrop equilibrium, Nash equilibrium, potential games, variational inequalities, pricing

1. Introduction

Determining the equilibrium state of a traffic network has been a preoccupation of transport planners for nearly half a century. Underlying this preoccupation is the assumption that road traffic will naturally arrange itself in an equilibrium flow, under steady state conditions. In this context, predictions of future traffic flow patterns that would follow any changes to the network or demand levels rely upon an accurate representation of the traffic equilibrium.

Except for recent years, telecommunication network flow models have not made use of the notion of network equilibrium, as traffic networks do. Indeed, while road traffic is highly individualized, each driver making his or her own route choice decision, telecommunication networks are much more centralized.

Recently, however, equilibrium models have started to emerge in telecommunication networks. This is due to two main reasons. First, the deregulation and privatization of large telecom companies introduced competitive and decentralized behavior among telecom operators. Secondly, a new concept of networks has been developed and deployed with great success, in which most of the intelligence lies at the edges of the network (at the sources and destinations) allowing for increased speed and reduced overhead and costs, as well as the possibility for intelligent (and selfish) route or provider choice decisions. The Internet provides one example of such an environment.

1.1. Wardrop and Nash equilibria

The definition of the steady state equilibrium of a traffic network was put forth by J.G. Wardrop in his 1952 treatise (Wardrop, 1952) which provided two different definitions of traffic assignment concepts. The first is commonly referred to as the Wardrop, or traffic equilibrium, principle and, as we will show later, is a variant of Nash equilibrium for networks. It states that "*The journey time on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route*." Wardrop's second principle is what has become known as the system optimum principle, and states that "*The average journey time is a minimum*." While some point to the economist Pigou, who stated analogous principles in his 1920 *Economics of Welfare*, as the rightful originator of these ideas, they had not been applied to networks, and transportation networks in particular, until Wardrop's seminal work.

Analogous to Wardrop's first principle is the definition of the famous Nash equilibrium defined in J.F. Nash's 1950 doctoral thesis. Expressing it in terms of network flows, *a flow pattern is in Nash equilibrium if no individual decision maker on the network can change to a less costly strategy, or, route*. In other words, the Nash equilibrium does not state what can or cannot happen when more than one decision maker changes their strategy (route) simultaneously. When the decision makers in a Nash game are discrete and finite in number, a Nash equilibrium can be achieved without the costs of all used routes being equal, contrary to Wardrop's equilibrium principle. In some cases, Wardrop's principle represents a limiting case of the Nash equilibrium principle, as the number of users becomes very large. We will present some such precise statements on this relationship later in this paper.

The second Wardrop principle, that of system optimality, assumes that congested networks can be globally optimized. While this can be true for a network which is entirely controlled by a single operator, it is not so with networks of road traffic or with disaggregated telecommunication networks. Therefore, in the remainder of this paper, we will focus on equilibria, rather than pure system optima.

1.2. Equilibria in networks

Given the fundamental nature of equilibria in many large-scale systems, it is of no surprise that researchers studying transportation networks have been preoccupied with developing models that reproduce this equilibrium, as a function of network characteristics and user demand levels. Typically, transport equilibrium models consider vehicles to be the fundamental units seeking an equilibrium, or, in the case of public transport, the individual traveler. In both of these cases, since the number of users is generally very large, the Wardrop concept, that treats individual user contributions to the costs as infinitesimal, is preferred to the (in this respect, more general) Nash paradigm.

In the context of telecommunication networks, the Wardrop equilibrium is used most often to model the situation in which the routed entities are packets, and routing decisions are taken at the nodes of the networks (rather than by the users) so as to minimize the (per-packet) delay. In many actual networks, the routers at the nodes seek to minimize the per-packet delay in terms of the number of "hops," or nodes, to the destination. There are, however, situations in which it is more advantageous to work with actual delays as

cost metrics, rather than the number of hops (see Bertsekas and Gallager, 1987; Gupta and Kumar, 1997), and it is in these cases that the Wardrop equilibrium has been used to describe the resulting flow patterns. This is the case, for example, in ad-hoc networks in which both users as well as base stations are mobile, or where there are no base stations so that users are responsible to relay messages of other users. For these type of networks a Wardrop type equilibrium has been advocated in Gupta and Kumar (1997).

Wardrop equilibria have also been used in telecommunication networks to model a large number of users that can determine individually their route and in which the routed object is a whole session, see Korilis and Orda (1999) (whose model includes in addition some side constraints on the quality of service).

A third context in which Wardrop equilibrium has been used outside of transportation is in distributed computer networks, in which the routed objects are jobs. An individual job can be processed in any of several interconnected nodes (computers) and the routing decision is taken so as to minimize its expected delay in the system (composed of both communication as well as processing delay). Much material on that application can be found in Kameda et al. (1997).

The most widely studied concept of equilibrium in telecommunication networks is, however, the Nash equilibrium applied to the case of *finitely* many decision makers. The decision makers typically represent service providers, each of whom can determine how to route the flow generated by its subscribers (and how to split flow between actual paths).

The paper is structured as follows. In Section 2.1, we describe the links between Wardrop and Nash equilibria and the classes of potential and population games. Since these sets are similar but not not equivalent, we hope that the description in this paper encourages transfers of results from one class of games to another. Section 2.3 discusses commodity-link flow models, which, along with the models of Section 2.4, are described by variational inequalities. Since the latter class of problems has been studied extensively in the transportation equilibrium literature, many of the developments could potentially be applied to models of telecommunication networks. In Section 2.5, we present a few non-additive cost equilibrium models; algorithmic work for this class of models could therefore serve both the transportation and telecom communities.

In Section 3, we present two forms of network pricing. First, in Section 3.1, we describe models in which link prices are given by Lagrange multipliers, and as such are a form of marginal cost pricing, including recent extensions to that type of pricing that allow auxiliary objectives, such as revenue maximization, to be taken into account, albeit to a limited extent. We then present in Section 3.2 the more general, bilevel, network pricing problem, which allows operator revenue to be fully maximized, and discuss some algorithmic approaches to solving that very complex problem.

2. Equilibrium models spanning transport and telecom

2.1. *Basic Wardrop model*

The basic equilibrium model in traffic networks falls into the category of potential games with an infinite number of users. Indeed, the Wardrop equilibrium condition (the first

Wardrop principle, stated above), can be expressed mathematically to state that the flow on
every route r serving a commodity, or origin-destination (OD) pair, w, is either zero, or its
cost is equal to the minimum cost on that OD pair. Along with the fact that the cost on any
route serving an OD pair is at least as high as the minimum cost on that OD pair, and the
satisfaction of demand for each OD pair, we obtain the following system:

$$h_{wr}(c_{wr} - \pi_w) = 0, r \in R_w, w \in W, \tag{1}$$

$$c_{wr} - \pi_w \geq 0, r \in R_w, w \in W, \tag{2}$$

$$\sum_{r \in R_w} h_{wr} = d_w, w \in W \tag{3}$$

where h_{wr} is the flow on route $r \in R_w$, the set of routes joining node pair $w \in W$, the set of
origin-destination node-node pairs. The cost or delay on that route is c_{wr}, and π_w is the min-
imum cost on any route joining node pair w. The demand for service between the node pair
w is denoted d_w. Define the directed graph as $G = (N, A)$ where N is the set of nodes and
A is the set of links. Assume that c_{wr} is the sum of link costs $t_l(x_l)$ of the links l along route
r and that the cost t_l is a function only of the total flow x_l over each link l. Then, adding non-
negativity restrictions $h_{wr} \geq 0$ and $\pi_w \geq 0$, the resulting system of equalities and inequal-
ities can be seen as the Karush-Kuhn-Tucker (KKT) optimality conditions of the following
optimization problem, known as the Beckmann transformation (Beckmann, 1956).

$$\min \quad f(x) = \sum_{l \in A} \int_0^{x_l} t_l(x_l) \, dx = \sum_{l \in A} \int_0^{\sum_{i \in N} x_{il}} t_l(x) \, dx$$

$$\text{subject to} \quad \sum_{r \in R_w} h_{wr} = d_w, w \in W, \tag{4}$$

$$\sum_{w \in W} \sum_{r \in R_w} h_{wr} \delta_{wr}^l = x_l, l \in A, \tag{5}$$

$$x_l \geq 0, l \in A, \tag{6}$$

where x_{il} is the flow of users from OD pair i on link l, x_l is $\sum_{i \in N} x_{il}$ and δ_{wr}^l is a $0 - 1$
indicator function that takes the value 1 when link l is present on route $r \in R_w$.

2.2. Wardrop equilibrium and potential games

The fact that Wardrop equilibrium can be obtained using an equivalent optimization problem
with a single player having some cost $f(x)$ is a feature common to a whole class of games
known as potential games. This class of games was formally introduced by Monderer and
Shapley (1996) for the case of finitely many players. It was extended in Sandholm (2000)
to the case of population games, which includes the setting of Wardrop equilibrium.

In developing the concept of potential games, game theorists seem not to have been aware
of the huge literature on road traffic equilibria starting from Wardrop (1952) and Beckmann
(1956). Monderer and Shapley (1996) write in: "To our knowledge, the first to use potential
functions for games in strategic form was Rosenthal (1973)." Interestingly enough, this
reference (see also Rosenthal, 1973b) includes a discrete version of Wardrop equilibrium
with finitely many players, called "congestion games."

The original definition of a potential for a game is as follows. Introduce the following N-player game $G = (N; (S^i)_{i \in N}; (\eta^i)_{i \in N})$ where S^i is the action set of player i, $S = \times_{i \in N} S^i$ and $\eta^i(s)$ is the payoff for player i when the multistrategy $s \in S$ is used. For $s \in S$, let $(s \mid t^i)$ denote the multistrategy in which player i uses t^i instead of s^i and other players $j \neq i$ use s^j. A potential for the game is defined in Monderer and Shapley (1996) as a real valued function P on S such that for each i, every $s \in S$ and every $t^i \in S^i$, $P(s \mid t^i) - P(s) = \eta^i(s \mid t^i) - \eta^i(s)$. Existence and uniqueness of equilibria of potential games in that setting has been established in Monderer and Shapley (1996) and Neyman (1997).[1]

An adaptation of this definition is needed for population games; see Chap. 3 of Patriksson (1994) and Sandholm (2000), in which there are N classes of populations of "infinitesimal" players, where the "mass" of players of type i is given by some constants d_i. Let $\alpha(j, t)$ be the fraction of members of population type j that use action $t \in S^i$. A multistrategy is the collection $\alpha = (\alpha(j, t))$. We assume that the payoff η^i for a player of class i is a function of his own action as well as of the multistrategy α. Let S^i be the set of actions available to a player of population i, $i = 1, \ldots, N$. We say that α^* is an equilibrium if, for any i, any $s \in S^i$ and any $t \in S^i$ such that $\alpha^*(i, t) > 0$, $\eta^i(t; \alpha^*) \geq \eta^i(s; \alpha^*)$. Equivalently, letting $f = -\eta$, we say that flow α^* is in equilibrium if the following variational inequality problem (VIP) holds for all $\alpha \in S$: $f(\alpha^*)(\alpha - \alpha^*) \geq 0$.

We then define P to be a potential for the population game if for each s, the vector of payoffs $\eta(s)$ is the gradient of $P(s)$. Under mild conditions on the payoff functions and strategy sets, one can thus establish the existence and uniqueness of equilibria in potential population games. (See Chap. 3 of Patriksson, 1994.)

2.3. Commodity—link-variable models

In telecommunication network models, the variable of interest is often the commodity-link flow, x_{il}, rather than the total link flow, x_l. In this section, we consider a typical model based on commodity-link flows as the decision variable, and through it, derive some relations between Nash and Wardrop equilibria.

We have seen that the Wardrop equilibrium can be computed using an equivalent convex optimization problem (related to the potential). We show below that reminiscence of this potential appears also in problems of Nash equilibria with finitely many users: all equilibria satisfying some conditions on the commodity-link flow (condition (11) below) have the same total link flow that can be obtained from some convex optimization problem.

In the network routing problem, the decision variables x_{il} are restricted by the non-negativity constraints for each link l and player i: $x_{il} \geq 0$ and by the conservation constraints for each player i and each node v, expressed below in terms of ingoing and outgoing nodes:

$$r_v^i + \sum_{j \in In(v)} x_{ij} = \sum_{j \in Out(v)} x_{ij} \tag{7}$$

where $r_v^i = d_i$ if v is the source node for player i, $r_v^i = -d_i$ if v is its destination node, and $r_v^i = 0$ otherwise; $In(v)$ and $Out(v)$ are respectively all ingoing and outgoing links of node v. (d_i is the total demand of player i).

A typical commodity-link cost function, for the ith commodity on link l, is given by

$$f_{il}(x_{il}) = x_{il}t_l\left(\sum_{j\in\mathcal{I}_l} x_{jl}\right), \tag{8}$$

where $\mathcal{I}_l$ is the set of commodities (players) using link l. Note that this model resembles the system optimal model defined in Wardrop's second principle. The Lagrangian with respect to the conservation of flow constraints is

$$L_i = \sum_{l\in\mathcal{A}} x_{il}t_l\left(\sum_{j\in\mathcal{I}_l} x_{il}\right) + \sum_{v\in\mathcal{N}} \pi_{i,v}\left(r_v^i + \sum_{j\in In(v)} x_{ij} - \sum_{j\in Out(v)} x_{ij}\right),$$

for each player i. Thus a vector x with nonnegative components satisfying (7) for all i and v is an equilibrium if and only if the following Karush-Kuhn-Tucker (KKT) condition holds:

Let $x_{uv} = \sum_{i\in\mathcal{A}} x_{i,uv}$, where $x_{i,uv}$ is the flow of users from OD pair i on the link defined by node pair u, v. There exist Lagrange multipliers $\pi_{i,u}$ for all nodes u and all players, i, such that for each pair of nodes u, v connected by a directed link (u, v),

$$t_{uv}(x_{uv}) + x_{i,uv}\frac{\partial t_{uv}(x_{uv})}{\partial x_{uv}} \geq \pi_{i,u} - \pi_{i,v},$$

with equality if $x_{i,uv} > 0$.

Define $\pi_u = \sum_i \pi_{i,u}$. Taking the sum over all players we get the following necessary conditions for x to be an equilibrium for each link (u, v):

$$It_{uv}(x_{uv}) + x_{uv}\frac{\partial t_{uv}(x_{uv})}{\partial x_{uv}} \geq \pi_u - \pi_v, \tag{9}$$

with equality if $x_{i,uv} > 0$ for all i, where I is the total number of players.

Assume that all players have the same source and destination, and let d be the sum of commodity demands. Then (9) are the KKT conditions for optimality of the vector $\{x_l\}$ with nonnegative components satisfying the conservation of flow constraints in the routing problem (single commodity) where the cost to be minimized is given by

$$\sum_{l\in A} x_l t_l(x_l) + (I - 1)\int_0^{x_l} t_l(y)\,dy, \tag{10}$$

and where the total demand to be shipped from the common source to the common destination is d; in particular, (9) holds with equality if $x_{uv} > 0$. Assume further that t_l are strictly convex, or more generally that expression (10) is strictly convex. Then this problem has a unique solution in total link flows, which we denote (x_l^*).

Now, let $\{x_{il}\}$ be a Nash equilibrium for the original problem having costs (8) with the property:

A1. Whenever $x_{il} > 0$ for some i and l then $x_{jl} > 0$ for all players j. (11)

A1 describes a property of the equilibrium: if (at equilibrium) one player sends positive flow through a link, then so do all other players. Under assumption A1, it follows that for all l, $\sum_i x_{il} = x_l^*$. Note however, that this is not true in general if A1 does not hold, since x_l^* need not be expressible as the sum over I of some nonnegative x_{il} that satisfy (7).

Remark. The above is an alternative proof to the one in Orda et al. (1993) of the uniqueness of the total link flows at all Nash equilibria satisfying A1.

Taking the limit in (10) as the number of players $I \rightarrow \infty$, the second term in (10) dominates, and by continuity of the functions and compactness of the feasible set, we observe that both the objective function and the solution approach that of the Wardrop equilibrium.

2.4. *General models and variational inequalities*

The basic equilibrium model presented at the end of Section 2.1 imposes a number of simplifications on the model of the traffic flow phenomenon, and in particular, on the travel time, or impedance, functions.

Most notably, for the potential function to exist, the travel time function, t_l, defined for each link of the network, $l \in A$, must be integrable. The most common way for this to occur is that the travel time on a link l depends only upon the flow present on the link l, that is, $t_l(x) = t_l(x_l)$. This simplification, in the traffic context, means that interactions between different traffic streams at junctions cannot be modeled within this paradigm (even if we use a virtual link to model the node), since then, the travel time on a link l that reaches the junction is a function of flows on some or all links meeting link l at the junction, that is, $t_l(x) = t_l(x_1, \ldots, x_l, \ldots, x_m)$ (and is not just a function of the sum of flows). The simplification also imposes that only a single class of users is modeled, since multiple classes of users would interact on each link, resulting once again in multivariate link travel time functions.

When the link travel time functions are multivariate, it is usually the case that no potential that can be obtained by integrating the travel time functions as in the basic model. Examples of multivariate link cost functions can be found in the literature on modeling signalized junctions on a road network, see Heydecker (1983). Other examples can be found in the modeling of multimodal networks, such as networks on which buses and cars share the road space or trucks and light vehicles, since each traffic class effects the traffic differently, and each class has its own travel time function, depending on all classes present on the link. Some characteristics of this type of multivariate cost functions can be found in Toint and Wynter (1995) and Marcotte and Wynter (2004).

Although the potential function approach cannot be used to describe the multivariate equilibrium, the Wardrop equilibrium conditions are valid regardless of whether the cost functions are univariate or multivariate, and they can be expressed for both types of cost functions in a compact variational framework.

In telecommunication network planning, link impedance functions can often be quite complex, due to the underlying probabilistic phenomena as well as the interacting cost components of delay, packet loss, jitter, etc... A typical form of the commodity-link cost functions (see, for example, Orda et al., 1993) is

$$t_{il}(x_{il}, x_l) = x_{il}t_l(x_l) = \frac{x_{il}}{C_l - x_l}, \tag{12}$$

where, as before, $x_l = \sum_{i=1..I} x_{il}$ is the flow of all classes i on link l. The constant C_l is the capacity of link l. The user classes in this case correspond to commodities, or

origin-destination demands. For more justification on this type of delay models, see Baskett et al. (1975), Kameda and Zhang (1995). In some simple settings, such as a network of parallel arcs, when the sum of the demands of all classes is less than link capacity, or on some one-commodity networks, Orda et al. (1993) show uniqueness of the Nash equilibrium. Also under "diagonal strict convexity," the authors show uniqueness of the Nash equilibrium, yet this condition only holds under quite restrictive conditions on the cost functions or on the topology (Altman et al., 2002). In Orda et al. (1993) and Altman and Kameda (2001), uniqueness of the link class flows is shown for costs of the form (12) for general networks under assumption A1 (see Eq. (11)).

We can analyze the *Wardrop* equilibrium for this system easily by expressing it as the solution $\{x_i^*\}$ of the following variational inequality:

$$t_i(x^*)^T (x_i^* - x_i) \le 0, \tag{13}$$

for all feasible class-flow vectors, x_i, for all i, where, as above, the vector $x = \sum_{i=1..I} x_i$ and $t_i = t$ for all classes $i = 1..I$, and determine, for example, if the Jacobian of the mapping t is singular, in which case we recover the nonuniqueness of the equilibrium in class-flow variables. Since the Jacobian of the mapping t is clearly singular ($\partial t_i(x)/\partial x_j = \partial t_j(x)/\partial x_i$ for all i, j), we recover the nonuniqueness of the equilibrium (in the variable x_i) that was observed by Orda et al. (1993) to occur on general networks.

We can also contrast the Wardrop and Nash equilibria through this example. The Nash equilibrium x_i^* satisfies

$$\sum_{l \in R_i} x_{il}^* t(x_{il}^*, x_{\neq il}^*) \le \sum_{l \in R_i} x_{il} t(x_{il}, x_{\neq il}^*)$$

for each user class $i = 1..I$, where the index $\neq i$ includes all classes not equal to i. Rewriting, we obtain that $\sum_{l \in R_i} x_{il}^* t(x_{il}^*, x_{\neq il}^*) - \sum_{l \in R_i} x_{il} t(x_{il}, x_{\neq il}^*) \le 0$ for each $i \in I$ and therefore when

$$t(x_{il}^*, x_{\neq il}^*) - t(x_{il}, x_{\neq il}^*) = 0, \tag{14}$$

the above reduces to $\sum_{l \in R_i} (x_{il}^* - x_{il}) t(x_{il}^*, x_{\neq il}^*) \le 0$, for each $i \in I$, which is equivalent to the (Wardrop) VIP with cost operator t and classes given by users $i \in I$. Indeed, (14) occurs precisely when the influence of an additional user on the cost, t, is 0.

Another example of general costs in telecommunications can be found in Altman et al. (2002), where the network model includes dropping of calls if capacity is exceeded. The cost criterion for each user class i is the probability that his message is rejected along its path, given by

$$B_i(x) = 1 - \frac{\sum_{\mathbf{m} \in S_i} \prod_{j=1}^{I} \left(x_j^{m_j}/m_j! \right)}{\sum_{\mathbf{m} \in S} \prod_{j=1}^{I} \left(x_j^{m_j}/m_j! \right)}.$$

S is the set of feasible "states"; S_i is the set of states for which another call of user i can still be accepted (without violating capacity constraints), $\mathbf{m}$ is the system state whose jth component m_j is the number of class j calls in the system. Hence the number of terms in the cost function depends upon the configuration and capacity constraints of the network. (The larger the number of feasible paths for a user class, the more terms present.) When each

class, which can be represented by an origin-destination pair, seeks a Wardrop equilibrium, the corresponding variational inequality is to find x_i^* such that $B_i(x^*)^T (x_i^* - x_i) \leq 0$, for all x_i, for all classes i. Even in the simplest topology of parallel links, it has been shown in Altman et al. (2002), that there may exist several equilibria with different total link flows.

Depending on the form of monotonicity satisfied by the cost operator in the variational inequality, one can choose convergent algorithms for its solution, as well as determine whether or not the solution will be unique. Weaker forms of monotonicity for which the mathematical properties and a number of convergent algorithms are known include pseudo-monotonicity and strong nested monotonicity. (See Marcotte and Wynter, 2004; Cohen and Chaplais, 1988, for the latter.)

2.5. *Additive versus non-additive models*

The most widely studied performance measure investigated to date in transportation, computer and telecommunication networks has been the expected delay. In transportation networks, this cost metric leads naturally to models in which the route costs are additive functions of link costs along the route, that is $c_r(x) = \sum_{l \in R} t_l(x_l)$.

In telecommunication networks, exogenous arrivals of jobs or of packets are often modeled as Poisson processes. Delays at links are modeled by infinite buffer queues with i.i.d. service times, independent of the interarrival times. In the particular case in which the service time at a queue is exponentially distributed, then whenever the input process has a Poisson distribution so will the output stream. This makes the modeling of service times through exponential distributions quite appealing, and makes the cost along a path of tandem queues additive. However, the expected delay turns out to be additive over constituent links in a general topology under a much more general setting, known as BCMP networks (named after its authors Baskett et al., 1975).

Indeed, as long as the exogenous arrivals have Poisson distributions and under fairly general assumptions on the service order and service distribution, the expected delay over each link is given by the expected service time divided by $(1-\rho)$, where ρ is the product of total average traffic flow at the queue and the expected service time. This general framework also allows for the modeling of multiclass systems, i.e. where different traffic classes require different expected service times at a queue, see Kameda and Zhang (1995).

Other more general types of separable additive cost functions have been used in telecommunication networks which can represent physical link costs due to congestion pricing, see e.g. Orda et al. (1993).

However, in both transportation and telecommunication networks, equilibrium models in which path costs are not the sum of link costs do arise.

In the transport sector, when environmental concerns are taken into account, such as the pollution associated with trips, non-additive terms arise. Path costs due to tolls or public transport costs are generally non-additive as well, since they are calculated over entire paths, and cannot be decomposed into a sum of the costs on component links.

In Gabriel and Bernstein (1997) and Bernstein and Wynter (2000), the authors discussed properties of a bicriteria equilibrium problem in which both time delay and prices are modeled on the links. Non-additivity arises from the nonlinear valuation of the tradeoff

between time and money, known as a nonlinear *value of time*. For example, users are willing to pay more (or less) per minute for longer trips than for shorter trips. That is, a route cost function may be expressed as $c_r(x, p) = V(\sum_{l \in R_i} t_l(x_l)) + \sum_{l \in R_i} p_l$, where $V : \Re^m_+ \mapsto \Re_+$ is the nonlinear value of time function.

In telecommunication networks, many important performance measures are neither additive nor separable. The first example is that of loss probabilities when the network contains finite buffer queues. We note that in this case there is no flow conservation at the nodes.[2] This model has been dealt with by light traffic approximations which are additive and separable, see Dinan et al. (2000) and Jiménez (2001).

Another performance measure (already mentioned in the previous subsection) that has been studied in the context of network equilibrium is that of rejection probabilities. The network consists of resources at each link, and requests for connections between a source and a destination. The resources are limited, and a connection can only be established if there are sufficiently many resources along each link of a route between the source and destination. For the case where connections arrive according to a Poisson process and where calls last for an exponentially distributed duration (these assumptions model telephone networks well), simple expressions for rejection probabilities (i.e. the probability that an arriving call will find the line busy) are available. These expressions are neither separable nor additive. Such networks have been studied in Altman et al. (2002) and Bean et al. (1997).

When route costs are no longer the sum of constituent link costs, algorithms for solving the network equilibrium problems must be modified; indeed, underlying most algorithms for the network equilibrium problem is a shortest path search, and standard searches all suppose additive path costs. One algorithm for the non-additive route cost model can be found in Gabriel and Bernstein (2000) while another is provided in the article by M. Patriksson, in this special issue.

3.   Pricing

The pricing of network services, in transportation and telecommunications, has taken on more importance in the past decade. In transportation networks, congestion pricing, such as road tolls which adapt to time of day, and could, in principle, vary with the congestion level, have become a reality in many parts of the world. Similarly, pricing of communications networks and the Internet have become intense areas of research activity since telecom deregulation, and due to the development of Internet services, such as tele-conferencing, software leasing, and e-commerce.

3.1.   Marginal cost pricing

The vast majority of the literature on pricing of networks supposes that the pricing scheme is based upon shadow prices, or marginal costs. That is, the marginal cost of adding a new request, or user, to the network defines the price that should be charged to that user. See Patriksson (1994) for a list of references from the transportation literature, and Kelly (1997), Kelly et al. (1998), and Low and Lapsley (1999) in the telecommunications literature.

When there are no capacity constraints in the model, the marginal cost of adding a new user on link $a \in A$ within the Wardrop equilibrium model is given by

$$t_a(x_a) + x_a t'_a(x_a). \tag{15}$$

The difference then between marginal costs and user costs is the second term, $x_a t'_a(x_a)$. For the user equilibrium solution to be "driven" to the system optimal solution, the cost of adding each new user must be taken into account, such as through marginal cost pricing. This means substituting delay function (15) for $t_a(x_a)$ in the Wardrop equilibrium model.

In telecommunication networks, capacity constraints are generally present on each link in the model. The Lagrange multipliers on those constraints provide another marginal cost, representing an additional cost that can be imposed on each user that would exceed the capacity on the link. By construction, the multipliers on those constraints, and therefore the link prices, are zero unless the link is at capacity. (Note that these prices need not, however, be unique.) This price can be understood as a control measure, a "cost" imposed in addition to the travel cost itself that keeps the flow within the boundaries of the capacity constraints, without actually having to enforce the capacity constraints, see e.g. Larsson and Patriksson (1999).

In the telecommunications literature, this notion of pricing is referred to as the Proportional Fairness principle, when the network flow problem is defined by the maximization of the sum of logarithmic user's utility functions, see Kelly (1997) and Kelly et al. (1998). That is, the network flow objective is given by $\max_{d \in D} \sum_{i \in I} \log(d_i)$ subject to flow conservation, link capacity, and non-negativity constraints, defined as the set D of feasible resource allocations, where $d \in \Re_+^W$ is the amount of bandwidth to be allocated to user $i \in I$, defined by an origin-destination node pair. The objective of the proportional fairness bandwidth allocation and pricing scheme is therefore to determine the origin-destination demand level to allocate to each pair that maximizes a logarithmic utility function, and satisfies network constraints.

In Larsson and Patriksson (1998) and Larson et al. (2002), the authors showed that the non-uniqueness of the Lagrange multipliers in capacity-constrained equilibrium problems could be used to optimize traffic management schemes by defining a secondary function over the set of multipliers. In Wynter (2001) and in Bouhtou et al. (2003), this technique was applied to the Proportional Fairness pricing model so as to permit revenue maximization using fair prices. That is, the network operator would solve $\max_{\lambda \in \Lambda(x^*)} \lambda^T x^*$, where x^* is a proportionally fair resource allocation and $\Lambda(x^*)$ is the (polyhedral) set of Lagrange multipliers at x^*. This approach has been studied and algorithms developed in Larson et al. (2002) and Rydergren (2001). In Bouhtou et al. (2003), numerical experience aimed at studying the degree to which revenue maximization is possible while still maintaining the properties of the marginal cost pricing scheme is provided.

3.2. *Pricing strategies with other objectives*

While Lagrange multiplier-based pricing strategies have a historical foundation, in that they quantify the notion of marginal cost pricing of networks, they do not take into account

fully the desire of the network operator to maximize revenue. Indeed, the link prices are heavily constrained by the fact that they must be Lagrange multipliers, or by the fact that they must precisely drive flows to the system optimal solution. The strategy described above does permit limited profit maximization to be achieved, when the total revenue is not unique, as it may be in some cases (see Larsson and Patriksson, 1998; Bouhtou et al., 2003).

However, in general, for revenue maximization to be realized, prices can no longer be tied to Lagrange multipliers. In the case of capacity constraints, the multipliers, as stated earlier, are only positive when capacity is reached. This alone reduces potential for revenue. Furthermore, in telecommunication networks, there are cases in which further integrality constraints are imposed: one cannot split the flows at any proportion but instead, one has to choose the transmission rate over each link among some given set. In such cases, capacity need not be reached exactly for the network to be congested. (See Jarray and Wynter, 2002, for a solution to this problem in a different context.)

The general paradigm that models optimal network pricing is known as the bilevel program; the lower level of the problem represents the user equilibrium or network flow problem with prices as a parameter, and the upper-level, or network operator's problem, in g, seeks to optimally set the prices given the response of the users or flow pattern, to those prices. The mathematical formulation is given by

$$\max g(x, p) = x^T p, \quad x \in S(p), \quad p \in P$$

where $S(p) = \{x^* : F(x^*, p)^T (x^* - x) \leq 0, x \in X(p)\}$, $X(p)$ being the set of constraints on flow variables x (network constraints, non-negativity, ... along with any other constraints that may depend on the price vector, p.) and where $F = \nabla f$ if the (Nash or Wardrop) equilibrium problem admits an equivalent convex optimization formulation. Recall that f are the cost functions of the players in the Nash game, or the objective function of the optimization form of the Wardrop equilibrium, in the separable case. The bilevel programming problem is known to be nonconvex and non-differentiable, even in the most favorable circumstances, when the users' equilibrium problem is constrained. It is therefore very difficult to solve over networks, and most methods in the literature are designed to converge to any stationary point of the problem. The most popular methods for solving the bilevel program to a local optimum include penalty methods, in which the constraint $x \in S(p)$, expressed by its KKT conditions, is penalized in the objective, g, as well as sensitivity-based methods, that obtain a subgradient of the objective, g, through sensitivity analysis and apply subgradient or bundle methods. (See for example Bard, 1998, and Patriksson, 2001, for the sensitivity-based methods.)

4. Perspectives

This paper has highlighted a number of areas in which common features between transportation and telecommunication network models exist, and where future research effort would be worthwhile. Here we summarize a few of the areas that seem particularly worthwhile for further research effort.

4.1. Different definitions of equilibrium and games

It is clear that through lack of a common forum, much work on network games is duplicated in different communities, often with the strength of the results varying across the board. We discussed some examples of this in Section 2.2, mentioning in particular that Wardrop equilibrium is a special case of a potential game, which is in turn a special case of population games. On the one hand, it is worth studying whether any results on population games generalize those known for Wardrop equilibria, or vice-versa. Most work on potential games considers a finite number of players, and so appear less useful in transportation networks than in telecommunications settings. We note that some work in discrete Wardrop equilibria can be found in Bernstein (1990), and some extensions of potential games to infinite many players can be found in Sandholm (2000).

4.2. General costs

We presented several settings in which costs are more general than the standard separable, total link flow cost functions so prevalent in the transportation literature. In particular, telecommunication models generally make use of commodity-link flow variables. While this can be seen as a special case of the multi-class models found in the transportation literature, we saw that through the particular form of the commodity-link costs, such as (12), a number of much stronger properties can be obtained. The most surprising of these is that the commodity-link flows can in some cases be unique (e.g., for networks of parallel links, ...). While this model and the uniqueness result does not immediately transfer to transportation networks, since the costs there will generally not satisfy the properties assumed, one could consider looking at particular networks and cost forms in transportation models to deduce much stronger properties than have been assumed to hold.

Similarly, by formulating more general models from telecommunications as variational inequalities, it is possible to determine their uniqueness properties over general networks quite readily, and, especially, to propose a much wider range of algorithms for solving for those equilibria.

Another non-standard form of equilibrium models are those which do not have additive costs over constituent links of the paths. Some work on these models exists in both communities, and we have cited a few references on such work. It seems clear that there should be considerable possibilities for adapting the results mentioned to the applications of both transport and telecom models.

4.3. Pricing

We have provided quite a cursory description of optimal network pricing, since a large body of literature on this topic exists in the telecommunication setting. However, the vast majority of that work does not include equilibrium models, or does not appear to be readily transferable to transportation modeling. We have therefore chosen not to include those references in this paper. We do, however, describe a well-known approach proposed for telecommunications pricing that can be shown to be a special case of an approach proposed

for transportation system management. The paradigm of bilevel programming seems particularly appropriate to both telecommunications and transportation pricing, when marginal costs are not sought. Adapting this paradigm to particular settings and applications will likely become a very active field of research. However, the algorithmic issues that will result are considerable; attention could be focused on devising more rapid and decomposable approximation algorithms.

Acknowledgments

Notes

1. Uniqueness is in fact established among the class of correlated equilibria, of which Nash equilibria is a subset.
2. Note that flow conservation fails even in the case of infinite queues when one considers multicast applications in which packets are duplicated at some nodes, see Boulogne and Altman (2002).

References

Altman, E., T. Başar, T. Jiménez, and N. Shimkin. (2002). "Competitive Routing in Networks with Polynomial Cost." *IEEE Trans. on Automatic Control* 47, 92–96.

Altman, E., R. El-Azouzi, and V. Abramov. (2002). "Non-Cooperative Routing in Loss Networks." *Performance Evaluation* 49(1–4), 43–55.

Altman, E. and H. Kameda. (2001). "Equilibria for Multiclass Routing Problems in Multi-Agent Networks." *40th IEEE Conference on Decision and Control*. Orlando, Florida, U.S.A.

Bard, J.F. (1998). "Practical Bilevel Optimization: Algorithms and Applications." Kluwer Book Series: Nonconvex Optimization and its Applications, Vol. 30.

Baskett, F., K.M. Chandy, R.R. Muntz, and F. Palacios. (1975). "Open, Closed, and Mixed Networks of Queues with Different Classes of Customers." *J. ACM* 22, 248–260.

Bean, N.G., F.P. Kelly, and P.G. Taylor. (1997). "Braess's Paradox in a Loss Network." *J. Appl. Prob.* 34, 155–159.

Beckmann, M., C.B. McGuire, and C.B. Winsten. (1956.) *Studies in the Economics of Transportation*. New Haven: Yale Univ. Press.

Bernstein, D.H. (1990)."Programmability of Continuous and Discrete Network Equilibria." PhD Thesis, University of Pennsylvania.

Bernstein, D.H. and L. Wynter. (2000). "Issues of Uniqueness and Convexity in Non-Additive Bi-Criteria Equilibrium Models." In *Proceedings of the EURO Working Group on Transportation*, Rome.

Bertsekas, D. and R. Gallager. (1987). *Data Networks*. NJ: Prentice Hall, Englewood Cliffs.

Bouhtou, M., M. Diallo, and L. Wynter. (2003). "A Numerical Study of Lagrange-Multiplier-Based Pricing Schemes." In P.M. Pardalos, I. Tsevendorj, and R. Enkhbat (eds.), *Optimization and Optimal Control*. World Scientific Publishing Co. (to appear).

Boulogne, T. and E. Altman. (2002). "Competitive Routing in Multicast Communications." Submitted. Available at http://www-sop.inria.fr/mistral/personnel/Thomas.Boulogne/

Cohen, G. and F. Chaplais. (1988). "Nested Monotony for Variational Inequalities over Product of Spaces and Convergence of Iterative Algorithms." *Journal of Optimization Theory and Applications* 59(3).

Dinan, E., D.O. Awduche, and B. Jabbarie. (2000). "Optimal Traffic Partitioning in MPLS Networks." In *Proceedings of IFIP-TC6 Networking*. Paris, France, May 14–19.

Gabriel, S.A. and D. Bernstein. (1997). "The Traffic Equilibrium Problem with Nonadditive Costs." *Transportation Science* 31, 337–348.

Gabriel, S.A. and D. Bernstein. (2000). "Nonadditive Shortest Paths: Subproblems in Multi-Agent Competitive Network Models." *Computational and Mathematical Organization Theory* 6, 29–45.

Gupta, P. and P.R. Kumar. (1997). "A System and Traffic Dependent Adaptive Routing Algorithm for Ad Hoc Networks." In *Proceedings of the 36th IEEE Conference on Decision and Control*. San Diego, pp. 2375–2380.

Heydecker, B.G. (1983). "Some Consequences of Detailed Junction Modeling in Road Traffic Assignment." *Transportation Science* 17, 263–281.

Jarray, F. and L. Wynter. (2002). "An Optimal Smart Market Auction for the Pricing of Telecommunications Services." INRIA Research Report 4310. Sixth INFORMS Telecommunications Conference, March 10–13, 2002, Hilton Deerfield Beach, Boca Raton, FL. Available at http://www.inria.fr/rrrt/index.en.html

Jiménez, T. (2001). "A Competitive Routing Approach for Packet Loss Minimization." (manuscript).

Kameda, H., J. Li, C. Kim, and Y. Zhang. (1997). *Optimal Load Balancing in Distributed Computer Systems*. Springer.

Kameda, H. and Y. Zhang. (1995). "Uniqueness of the Solution for Optimal Static Routing in Open BCMP Queueing Networks." *Math. Comput. Modelling* 22(10–12), 119–130.

Kelly, F. (1997). "Charging and Rate Control for Elastic Traffic." *European Trans. Telecommunications* 8, 33–37.

Kelly, F., A.K. Maulloo, and D.K.H. Tan. (1998). "Rate Control for Communication Networks: Shadow Prices, Proportional Fairness, and Stability." *Journal of the Operational Research Society* 49, 237–252.

Korilis, Y. A. and A. Orda. (1999). "Incentive-Compatible Pricing Strategies for QoS Routing." In *Proceedings of IEEE INFOCOM'99*. New York, NY, USA.

Larsson, T., J. Lundgren, M. Patriksson, and C. Rydergren. (2002). "Current Trends in Transportation and Network Analysis—Papers in Honor of Michael Florian." In Michel Gendreau and Patrice Marcotte (eds.), Kluwer Applied Optimization Series, Vol. 63, pp. 147–164.

Larsson, T. and M. Patriksson. (1998). "Side Constrained Traffic Equilibrium Models-Traffic Management Through Link Tolls." In P. Marcotte and S. Nguyen (eds.), *Equilibrium and Advanced Transportation Modelling*. Boston, MA: Kluwer Academic Publishers, pp. 125–151.

Larsson, T. and M. Patriksson. (1999). "Side Constrained Traffic Equilibrium Models: Analysis, Computation and Applications." *Transportation Research, Part B* 33, 233–264.

Low, S.H. and D.E. Lapsley. (1999). "Optimization Flow Control I: Basic Algorithm and Convergence." *IEEE/ACM Transactions on Networking* 7, 961–874.

Marcotte, P. and L. Wynter. (2004). "A New Look at the Multi-Class Network Equilibrium Problem." *Transportation Science* (to appear).

Monderer, D. and L.S. Shapley. (1996). "Potential Games." *Games and Econ. Behavior* 14, 124–143.

Neyman, A. (1997). "Correlated Equilibrium and Potential Games." *International Journal of Game Theory* 26, 223–227.

Orda, A., R. Rom, and N. Shimkin. (1993). "Competitive Routing in Multi-User Environments." *IEEE/ACM Trans. on Networking* 510–521.

Patriksson, M. (1994). *The Traffic Assignment Problem: Models and Methods*, VSP, Utrecht.

Patriksson, M. (1999). *Nonlinear Programming and Variational Inequalities: A Unified Approach*. Dordrecht: Kluwer Academic Publishers.

Rosenthal, R.W. (1973). "A Class of Games Possessing Pure Strategy Nash Equilibria." *Int. J. Game Theory* 2, 65–67.

Rosenthal, R.W. (1973). "The Network Equilibrium Problem in Integers." *Networks* 3, 53–59.

Rydergren, C. (2001). "Decision Support for Strategic Traffic Management: An Optimization-Based Methodology." PhD Dissertation, Department of Mathematics, Linköping University, Sweden.

Sandholm, W.H. (2000). "Potential Games with Continuous Player Sets" (submitted). See also "Evolutionary Justification of Nash Equilibrium" by the same author, PhD thesis, Northwestern University, 1998.

Toint, P. and L. Wynter. (1996). "Asymmetric Traffic Assignment, a Coherent Formulation." In J.-B. Lesort (ed.), *Proceedings of the 13th International Symposium on Transportation and Traffic Theory, (ISTTT)*. Lyon, France: Pergammon, pp. 237–260.

Wardrop, J.G. (1952). "Some Theoretical Aspects of Road Traffic Research." *Engineers Part II* 325–378.

Wynter, L. (2001). "Optimizing Proportionally Fair Prices." INRIA Research Report 4311. Available at http://www.inria.fr/rrrt/index.en.html