

A pricing methodology for resource allocation and routing in integrated-services networks with quality of service requirements

Tudor Mihai Stoenescu¹, Demosthenis Teneketzis²

¹University of Michigan, 1301 Beal Ave. Ann Arbor Mi. 48109-2122, USA
e-mail: stoenesc@eecs.umich.edu

²University of Michigan, 1301 Beal Ave. Ann Arbor Mi. 48109-2122, USA
e-mail: teneket@eecs.umich.edu

Manuscript received: December 2001/Final version received: May 2002

Abstract. We present an approach to the admission control, resource allocation and routing problem in connection-oriented networks that offer multiple services to users. Users' preferences are summarized by means of their utility functions, and each user is allowed to request more than one type of service. Each requested service may be delivered over one of many possible routes. Multiple types of resources are allocated at each link along the path of a connection. We assume that the relation between Quality of Service (QoS) and resource allocation is given, and we incorporate it as a constraint into a static optimization problem. The objective is to determine the amount of required resources and route for each type of service to maximize a welfare function that is equal to the sum of the users' utilities. We describe a competitive market economy that achieves the objective and satisfies the informational constraints imposed by the nature of the decentralized resource allocation and routing problem.

Key words: Pricing, Routing, Resource Allocation, Quality of Service, Integrated Services Networks

1 Introduction

Today's physical networks are supposed to support multiple heterogeneous applications, that is, applications with different traffic characteristics and different Quality of Service (QoS) requirements. Packet-switched, connection-oriented networks have been proposed to offer the QoS guarantees in integrated-services networks, because in connectionless networks individual packets may exhibit a significant variation in network service quality.

An interesting problem in integrated services connection oriented networks is the determination of resource allocation and routing schemes for general admission, satisfying the following constraints: (i) they efficiently allocate resources at each link, appropriately distributing the QoS among the various resources at each link, such that the QoS requirements for all accepted services are satisfied; (ii) they are social-welfare maximizing; (iii) they satisfy the informational constraints imposed by the network: the network is an informationally decentralized system where the number of users is unknown, users' preferences are private information, users are interested only on their requested services and are indifferent on the particular resource allocation schemes that satisfies the services.

To solve the aforementioned problem one must have: (a) a formal relation between resource allocation and QoS requirements; (b) a mechanism that allocates resources to individual users, satisfies the QoS requirements for each user, is social welfare maximizing, and satisfies the informational constraints imposed by the network.

The above considerations, in particular (iii) and (b), led to the use of microeconomic methods for the analysis of these type of problems. The two major microeconomic approaches generally used for development of efficient decentralized resource allocation schemes in integrated service networks are resource-directed and price-directed, [6]. In the resource-directed approach, each user computes the marginal values for his current resources, and communicates them to the rest of the users. The allocation is then changed so that users with an above average marginal utility receive more of this resource and users with a below average marginal utility receive less. This approach has been used in [11] to develop decentralized algorithms for optimally allocating a single resource to a set of interconnected computing agents. In the price-directed approach, an initial allocation of resources is made and an arbitrary set of systemwide initial resource prices is chosen. Prices then are iteratively changed to accommodate the "demands" for resources until the total demand for a resource exactly equals the total amount available. Most of the results on decentralized resource allocation currently available in the literature are based on the price-directed approach [2–5, 7–10, 12, 14, 15, 17–20, 22, 25].

The work currently available on decentralized resource allocation by price-directed methods has addressed, either by analysis [3–5, 7, 9, 10, 12, 14, 18, 22, 25], or simulation and analysis [2, 5, 17, 19, 20], a subset of the issues outlined in the second paragraph of this section. A significant part of this work has dealt with single link networks [4, 17, 20, 25], or with the allocation of a single resource per connection [4, 5, 7, 10, 14, 17, 18, 20, 25].

In this paper we have adopted a philosophy, that appears in several papers [4, 5, 7, 10, 12, 18, 23, 25], to find a solution to resource allocation and routing by price-directed methods. This philosophy/methodology consists of the following steps: (1) formulate a centralized constrained optimization problem where the objective is the maximization of a social welfare function satisfying the constraints imposed by the QoS requirements and the availability of network resources; (2) use pricing methods to devise a decentralized resource allocation and routing scheme that realizes the solution of the centralized problem and satisfies the informational constraints imposed by the network. The existence of a solution to the centralized problem is shown, and market methods are used to structure and develop the solution. The existence of a set of prices that induce users to request the "optimal" allocation and

routing is established and, in some cases [4, 5, 12, 23, 25], an iterative scheme for adjusting the prices based on users' requests is described. However, none of the papers specify a mechanism to force the successive prices to converge to the "optimal" set of prices.

In this paper we find a solution to the admission control, resource allocation and routing problem for integrated service networks, by following the price-directed approach and the philosophy discussed in the previous paragraph. We consider a connection oriented network, on which users are allowed to request multiple services. Each service may be delivered along one of many potential routes. Users' preferences, summarized by their utility functions, are unknown to the network. We assume that, for each particular service, there is a known relation between the resource allocation and the QoS requirement, and we include this as a constraint in a static optimization problem. The objective of this optimization problem is to determine the route and the amount of the required resources for each service, in order to maximize the users' total welfare. We prove that a solution for this problem exists, and we describe a competitive market economy that implements the solution while satisfying the informational constraints imposed by the nature of the network. The economy considered, consists of four different types of agents: resource providers, service providers, users, and an auctioneer. The resource providers sell resources to the service providers. The service providers form services using these resources, and deliver these services to the users. The job of the auctioneer is to regulate the prices of the resources based on the observed aggregate excess demand. We specify an iterative procedure that is used by the auctioneer to update the prices, and we prove its convergence. Our philosophy and approach are similar to [23], but our model is more general than that of [23]. The specific contributions of this paper and its comparison with [23] are discussed in Section 2 after the mathematical model is specified and the resource allocation and routing problem is formulated.

The remainder of the paper is organized as follows: In Section 2 we formulate a centralized optimal resource allocation and routing problem. In Section 3 we prove the existence of a solution for this problem. In Section 4 we describe (Section 4.1) and analyze (Section 4.3) a competitive market economy that leads to a decentralized allocation and routing scheme that achieves a solution of the centralized optimal resource allocation and routing problem. In Section 5 we summarize the results, we discuss assumptions made and the properties of our solution.

2 Problem formulation

We first start by defining some of our notation. We let $\mathbf{L} = \{1, 2, \dots, L\}$ be the set of links in the network. The network provides one way connections to a set $\mathbf{N} = \{1, 2, \dots, N\}$ of users. Each user can request a set $\mathbf{M}_i = \{1, 2, \dots, M_i\}$ of *types* of connections. The connection types are characterized by the origin, destination and the quality of service (QoS) the network is required to provide once the connection is accepted. We let $\mathbf{T}^{i,j}$ be the set of possible routes that can be used for user i for the connection of type j . We assume that for all (i, j) , $\mathbf{T}^{i,j}$ is finite. We denote by $x^i = (x_1^i, x_2^i, \dots, x_n^i) \in \mathbb{R}_+^{M_i}$ the request vector of user i . The preferences of i in $\mathbb{R}_+^{M_i}$ are defined by the quasi-linear utility function $x_0^i + u_i(x^i)$, where $x_0^i > 0$ is the numeraire commodity [16].

We consider the following problem. During the call setup, the users announce to the network their requests. We denote by $\mathbf{V}^{i,j,t}$ the set of links that belong to the route $t \in \mathbf{T}^{i,j}$. The network has to determine the number of connections that will be accepted, the route each connection of each user will take, as well as the amount of resources that have to be allocated along each route in order to guarantee the QoS. The criterion for the allocation of resources is the maximization of the social welfare function $\sum_{i \in \mathbf{N}} u_i(x^i)$, which is the total utility of all the users.

We denote by the set $\mathbf{K} = \{1, 2, \dots, k\}$ the set of *types* of resources available to the network. We denote $r_{l,k}^{i,j,t}$ be the amount of type k resource that is reserved at link l along the route $t \in \mathbf{T}^{i,j}$ ¹. Along link l the maximum amount of resource k is denoted by $R_{l,k} > 0$. For every $t \in \mathbf{T}^{i,j}$ we denote $\mathbf{F}^{i,j,t} = \{r^{i,j,t} \triangleq r_{l,k}^{i,j,t}, l \in \mathbf{L}, k \in \mathbf{K}\}$ the set of resource allocations that guarantee the QoS for connection type j for user i .

The sets $\mathbf{F}^{i,j,t}$, which we assume to be known for every (i, j, t) , result in from relations that describe the interaction between resource allocation and QoS requirements along routes. Such relations can be found, for example, in [22] (for wired networks) and in [24] (for single-hop wireless networks).

We make the following assumptions on the utility functions u_i and the sets $\mathbf{F}^{i,j,t}$. We discuss these assumptions later in Section 5.

- A1:** The functions u_i are continuous, differentiable, locally non-satiated² and strictly concave in their argument, for all $i \in \mathbf{N}$.
A2: The sets $\mathbf{F}^{i,j,t}$, $t \in \mathbf{T}^{i,j}$ are compact³.

Based on the above discussion and assumptions we formulate the following optimization problem:

$$\max_{x, r, \mathbf{T}^{i,j}} \sum_{i \in \mathbf{N}} u_i(x^i) \quad \text{Max 1}$$

subject to the constraints:

$$x^i \in \mathbf{X}^i \triangleq \{x^i \in \mathbb{R}_+^{\mathbf{M}_i} : 0 \leq x_j^i \leq B < \infty, j \in \mathbf{M}_i\}, \quad \forall i \in \mathbf{N} \quad (2.1)$$

$$r^{i,j,t} \in \mathbf{F}^{i,j,t}, \quad t \in \mathbf{T}^{i,j}, \quad i \in \mathbf{N}, \quad j \in \mathbf{M}_i \quad (2.2)$$

$$\sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{M}_i} x_j^i r_{l,k}^{i,j,t} \leq R_{l,k}, \quad l \in \mathbf{L}, \quad k \in \mathbf{K} \quad (2.3)$$

The salient features of the above problem are:

- (F1)** Each user can request multiple services. Each service is characterized by its own QoS requirements.

¹ We assume that $r_{l,k}^{i,j,t} = 0$ if link l does not belong to the path $\mathbf{V}^{i,j,t}$, assigned to the connection.

² That is, $\forall x \in \mathbf{X}$ and $\varepsilon > 0$, $\exists x' \in \mathbf{X}$ such that $\|x' - x\| \leq \varepsilon$ and $u_i(x') > u_i(x)$.

³ We will see in Section 3, that compactness is the necessary condition for the existence of an optimal allocation.

- (F2) There are multiple possible routes, each consisting of multiple links, that can be chosen for each connection type j of each user i .
- (F3) The QoS requirements for each service are satisfied by appropriate allocation of resources along each path (route) available for that service. Such allocations of resources appear as constraints on the optimization problem.
- (F4) Any service for any user can not be split among different routes.

The model presented and the optimization problem formulated in this paper are inspired by the work in [23]. The contribution of this paper is two-fold:

- (i) The simultaneous solution of the resource allocation and routing problems within the context of Problem **Max 1**. Previously, in [23], only the resource allocation problem was solved within the context of Problem **Max 1**.
- (ii) The relaxation of the strict convexity constraint previously imposed (see [23]) on the sets $\mathbf{F}^{i,j}$. The sets $\mathbf{F}^{i,j}$, resulting from the relationship between resource allocation and QoS requirements, are usually compact but not convex (see [22]). Previous work, [23], which required the sets of allowable allocations to be strictly convex, had to consider only the allocations contained in the largest strictly convex subset of $\mathbf{F}^{i,j}$. This restricted significantly the sets of allowable allocations, thus resulting in a significant degradation in network performance. The removal of the requirement on the convexity of the sets $\mathbf{F}^{i,j}$ allows us to consider all possible resource allocations that satisfy the QoS requirements of different services. Therefore, the optimal allocations determined under the constraints of this paper are superior to those of [23] (even when routing alternatives are not considered in our problem formulation).

3 Existence of a welfare-maximizing solution

The main result of this section is summarized by the following theorem:

Theorem 1. *There exists a solution to problem **Max 1**.*

Proof. We let $\mathcal{T} \triangleq \{(t_{i,j}): \forall i \in \mathbf{N}, j \in \mathbf{M}_i, \text{ then } t_{i,j} \in \mathbf{T}^{i,j}\}$. So each element of \mathcal{T} is a set of routes, with a unique route corresponding to each (i, j) pair. Choose an element of \mathcal{T} , say t , and fix it. In [23] it has been proven that that for fixed t there exists a solution to **Max 1**. Since \mathcal{T} is finite, the solution to **Max 1** is the maximum of a finite set of solutions. This completes the proof of Theorem 1.

4 Market interpretation

In this section we describe an algorithm that converges to a solution of Problem **Max 1**, and satisfies the informational constraints imposed by the nature of the network problem. We proceed as follows: First we describe a competitive market economy consisting of resource providers, service providers, users and an auctioneer. Within the context of this market we specify a procedure,

used by the auctioneer, which leads to a resource allocation that achieves a solution of Problem **Max 1**.

4.1 Description of the market

The economy consists of the following four types of agents: resource providers, service providers, users and an auctioneer. The resource providers, service providers and users are price takers. They act as if their behavior has no effect on the equilibrium prices reached by the market allocation process. In our market the resources at each link are sold as raw material to the service providers. The price for resource of type k at link l will be denoted by $\lambda_{l,k}$. The service providers buy the resources from the resource providers. Using these resources, they set up services and the corresponding prices for each unit of these services. Then, they sell these services to the users.

We make the following observations: Because of the informational constraints mentioned above, we have two markets, one between the resource providers and the service providers, and the other between the service providers and the users. The price taking assumption and the fact that we try to maximize the users utilities imply that: (i) the service providers will not attempt to make a profit; and (ii) the service prices are directly derived from the resource prices. Therefore, we can look at the two markets as being one market.

4.1.1 Resource providers: We assume that resource providers own the resources at each link, and that there is no cost associated with the supply of the resources to the market. Let

$$y(\lambda) \in \arg \max_{y \in Y} \sum_{l \in \mathbf{L}} \sum_{k \in \mathbf{K}} \lambda_{l,k} y_{l,k} \quad (4.1)$$

be the aggregate supply at price $\lambda \triangleq \{\lambda_{l,k}\}_{l \in \mathbf{L}, k \in \mathbf{K}}$, with $Y \triangleq \{y \in \mathbb{R}^{KL} : 0 \leq y_{l,k} \leq R_{l,k}, l \in \mathbf{L}, k \in \mathbf{K}\}$. From the assumption that there is no cost associated with the supply of resources to the market, equation (4.1) has the solution $y_{l,k} = R_{l,k}$ for all $l \in \mathbf{L}$, $k \in \mathbf{K}$ and $\lambda \geq 0$.

4.1.2 Service providers: The users request services from the service providers. Each of these requests is described by the origin, destination and the minimal level of quality required. The services are indexed by the (i, j) pair, with $i \in \mathbf{N}$ representing the user and $j \in \mathbf{M}_i$ representing the service type. For each pair (i, j) there exists a set $\mathbf{T}^{i,j}$ of possible routes that can be used. The service providers allocate resources $r^{i,j,t}(\lambda) \in \mathbf{V}^{i,j,t}$, with $t \in \mathbf{T}^{i,j}$, so that the minimum cost for the service and the lowest acceptable quality level are attained. Through the rest of the paper we assume that each service can not be distributed over multiple routes. We shall show that due to the pricing method, that we will describe later, this assumption does not change the nature of our solution.

The assumption that the service providers are price takers and that they are freely entering and exiting the market, implies that the service providers are not profit makers. So service providers allocate the resources for each type of connection by solving:

$$r^{i,j,t}(\lambda) \in \arg \min_{r^{i,j,t} \in \mathbf{F}^{i,j,t}} \sum_{l \in \mathbf{V}^{i,j,t}} \sum_{k \in \mathbf{K}} \lambda_{l,k} r_{l,k}^{i,j,t} \quad (4.2)$$

where $i \in \mathbf{N}$, $j \in \mathbf{M}_i$, $t \in \mathbf{T}^{i,j}$. For each (i, j) pair, equation (4.2) generates a set of routes along which the price per unit of service is minimal. If for some (i, j) there are more than one route of minimum price, the service provider picks one of these routes. A more detailed discussion on how the provider picks this route is presented in section 5.

Based on the assumptions above, and the resources allocated for the services, the following price per unit of service is announced to the users:

$$p_j^i(\lambda) \triangleq \sum_{l \in \mathbf{V}^{i,j,t}} \sum_{k \in \mathbf{K}} \lambda_{l,k} r_{l,k}^{i,j,t}, \quad i \in \mathbf{N}, j \in \mathbf{M}_i \quad (4.3)$$

where $r_{l,k}^{i,j,t}$ are determined by (4.2).

4.1.3 Users: Users request one way connections from the service providers. Based on the price $p_j^i(\lambda)$, announced by the service providers, the users demand a number of connections determined by

$$x^i(p) \in \arg \max_{x^i \in \mathbf{X}^i} \left[u_i(x^i) - \sum_{j \in \mathbf{M}_i} x_j^i p_j^i(\lambda) \right], \quad \forall i \in \mathbf{N} \quad (4.4)$$

4.1.4 Auctioneer: The role of the auctioneer is to regulate the prices of the resources. He does this based on the aggregate excess demand vector $z(\lambda, \mathbf{t})$:

$$z_{l,k}(\lambda, \mathbf{t}) \triangleq \sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{M}_i} (x_j^i(\lambda) r_{l,k}^{i,j,t}(\lambda)) - y_{l,k}(\lambda) \quad (4.5)$$

where $l \in \mathbf{L}$, $k \in \mathbf{K}$ and $\mathbf{t} \in \mathcal{T}$ is determined by (4.2).

4.2 The market mechanism

In this section we present a market mechanism, described by an algorithm, called Algorithm (Q), that describes how the market works. The algorithm proceeds iteratively as follows:

- Step 1:** The auctioneer announces prices λ for the resources at each node of the network. The users announce their desired services to the service provider.
- Step 2:** Based on the auctioneer's announcement, the service provider computes the minimum price per unit of service according to (4.2) and (4.3). The service provider announces these prices to the users.
- Step 3:** Based on the prices p announced by the service provider, the users request services in the amount $x(p)$ satisfying (4.4).
- Step 4:** Based on the service demand vector $x(p)$, the auctioneer computes, through (4.5) the excess demand vector $z(\lambda)$.
- Step 5:** If $z(\lambda) \leq 0$ then the process ends. Otherwise the auctioneer changes the prices λ of resources according to a specific mechanism, announces new prices, say λ' , and the process is repeated from Step 2 on.

A detailed description of the price adjustment mechanism used by the auctioneer is presented in Appendix A. In Appendix B, we present a flowchart of the Algorithm **(Q)** from the point of view of the auctioneer. In Section 4.3 we prove this mechanism eventually leads to a resource allocation that achieves a solution of Problem **Max 1**. Consequently, the algorithm described in this section “approximates” in a finite number of steps an optimal solution to the original resource allocation problem (Problem **Max 1**) and satisfies the informational constraints imposed by the decentralization of information at the network.

4.3 Analysis of the market

The main result proved in the section is summarized by the following theorem:

Theorem 2. *There exists a price adjustment mechanism for the auctioneer such that algorithm **Q** leads to an allocation that achieves an optimal solution for Problem **Max 1**.*

The proof of Theorem 2 proceeds in several steps. First, we discuss the main ideas behind the proof. Then, we present preliminary technical results. Finally, we use the preliminary technical results to conclude the proof.

4.3.1 Discussion We prove Theorem 2 by presenting a specific price adjustment mechanism for the auctioneer that eventually leads to a resource allocation that solves Problem **Max 1**. The price adjustment mechanism is based on an algorithm by Scarf that determines the appropriate shadow prices for the constraints of **Max 1** and therefore the solution to the corresponding unconstrained optimization problem.

4.3.2 Preliminary results We proceed with the details of the analysis. Scarf’s Algorithm works in the price simplex. Therefore, we start by defining the following $KL + 1$ dimensional simplex.

$$S \triangleq \left\{ q \in \mathbb{R}_+^{LK+1} : \sum_{m=0}^{LK} q_m = 1 \right\} \quad (4.6)$$

where L and K are the cardinality of \mathbf{L} and \mathbf{K} respectively. For each $q \in S$ with $q_0 > 0$ we define the *price vector* λ' :

$$\lambda' = \{\lambda'_{l,k}\}_{l \in \mathbf{L}, k \in \mathbf{K}} \triangleq \left\{ \frac{q_1}{q_0}, \frac{q_2}{q_0}, \dots, \frac{q_{KL}}{q_0} \right\} \quad (4.7)$$

The goal is to find an optimal price for each of the resources in the network, that is, a price vector λ' which leads to a solution for **Max 1**. To achieve this goal we need to introduce the following concepts:

- (i) The subsets $P^D \triangleq \{q \in S : q = (\frac{n_0}{D}, \frac{n_1}{D}, \dots, \frac{n_{LK}}{D}), n_i \in \mathbb{N}\}$, where $D \in \mathbb{N}$ represents “how close” our solution is to the solution of **Max 1**. (In Sections 5 we discuss what we exactly mean by “how close”, and its implication to the problem.)
- (ii) The notions of a *side* of the simplex and of a *primitive set* that are defined as follows:

Definition 1. A side of S , denoted by s^m , is defined by $s^m \triangleq \{q \in S : q_m = 0, m = 0, 1, \dots, LK\}$.

Definition 2. Let $\mathbb{S} \triangleq \{s^0, \dots, s^{LK}\}$ be the set of sides of S . Define $S^D \triangleq P^D \cup \mathbb{S}$.

Definition 3. $Q = (s^{i_1}, \dots, s^{i_n}, q^{j_0}, \dots, q^{j_{KL-n}})$, in S^D is called a primitive set if $q^{j_0}, \dots, q^{j_{KL-n}} \in P^D$, $s^{i_1}, \dots, s^{i_n} \in \mathbb{S}$, and no $q \in P^D$ is interior to the simplex generated by the vectors of Q , i.e. $\{x \in S : x_{i_1}, \dots, x_{i_n} \geq 0, x_m \geq \min\{q_m^{j_0}, \dots, q_m^{j_{KL-n}}\}, \forall m \neq \{i_1, \dots, i_n\}\}$.

Let $q = \left(\frac{D-(LK+1)}{D}, \frac{1}{D}, \dots, \frac{1}{D}\right)$ and $\{s^2, \dots, s^{LK}\}$ represent the set of sides with the first one removed. It is easily seen that $\mathcal{Q} \triangleq \{q, s^2, \dots, s^{LK}\}$ is a primitive set in S^D and that it is the unique primitive set containing $\{s^2, \dots, s^{LK}\}$.

(iii) The following subsets of S :

$$C_0 \triangleq \{q \in S : q_0 = 0 \text{ or } z_{l,k}(\lambda', \mathbf{t}) \leq 0, \\ \forall l \in \mathbf{L}, k \in \mathbf{K} \text{ and some } \mathbf{t} \in \mathcal{T}\}, \quad (4.8)$$

$$C_{l,k} \triangleq \{q \in S : q_0 > 0 \text{ and } q_{(l-1)K+k} = 0 \\ \text{or } z_{l,k}(\lambda', \mathbf{t}) \geq 0, l \in \mathbf{L}, k \in \mathbf{K}, \mathbf{t} \in \mathcal{T}\}, \quad (4.9)$$

where,

$$z_{l,k}(\lambda, \mathbf{t}) = \sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{M}_i} x_j^i(\lambda) r_{l,k}^{i,j,t}(\lambda) - R_{l,k}, \quad (4.10)$$

with $l \in \mathbf{L}, k \in \mathbf{K}, \mathbf{t} \in \mathcal{T}$, and $r(\lambda), x(\lambda)$ being solutions of the problems:

$$\min_{i \in \mathbf{T}^{i,j}} \min_{r^{i,j,t} \in \mathbf{F}^{i,j,t}} \sum_{l \in \mathbf{L}} \sum_{k \in \mathbf{K}} \lambda_{l,k} r_{l,k}^{i,j,t}, \quad \forall i \in \mathbf{N}, j \in \mathbf{M}_i, \quad (4.11)$$

$$\max_{x^i \in \mathbf{X}^i} \left[u_i(x^i) - \sum_{j \in \mathbf{M}_i} x_j^i \sum_{l \in \mathbf{L}} \sum_{k \in \mathbf{K}} \lambda_{l,k} r_{l,k}^{i,j,t}(\lambda) \right], \quad \forall i \in \mathbf{N}. \quad (4.12)$$

(iv) The concept of a *labeling function* that is defined as follows:

Definition 4. A labeling function is a function with domain S and range $\{0, 1, \dots, LK\}$.

We define the labeling function \mathfrak{g} as follows:

$$\mathfrak{g}(q) = \begin{cases} i & \text{if } q = s_i \\ 0 & \text{if } q \in C_0 \\ j & \text{where } j = \min\{(l-1)K + k : q \in C_{l,k}\} \end{cases} \quad (4.13)$$

A key result in our analysis is the following:

Lemma 1. *Starting with S^D , if we give $s^i \in \mathbb{S}$ label i for every $i \in \{0, \dots, LK\}$ and every $q \in P^D$ a label from $\{0, \dots, LK\}$, then exists a primitive set in S^D such that its vectors have distinct labels.*

Proof. See Appendix C.

Lemma 1 can be used to prove the following result that is crucial in the proof of the Theorem 2.

Lemma 2. *Let $\{D_i\}_{i \in \mathbb{N}}$ be a sequence such that $\forall i, D_i \in \mathbb{N}$ and $D_i \rightarrow \infty$ as $i \rightarrow \infty$. Let for every i , W^{D_i} denote the primitive set with distinct labels of S^D described by Lemma 1. Then there exists a subsequence $\{D'_j, j \in \mathbb{N}\}$ such that for any $q_{D'_j} \in W^{D'_j}$, $q_{D'_j}$ converges to λ' , where λ' solves Problem **Max 1**.*

Proof. First we note that given any $D \in \mathbb{N}$, the function \mathcal{G} satisfies the conditions for the labeling of the vectors S^D described by Lemma 1. This implies that given any $D \in \mathbb{N}$, there is a primitive set W^D with all the vectors that generate it having a distinct labels.

Define

$$C \triangleq C_0 \cap \left(\bigcap_{l \in \mathbf{L}, k \in \mathbf{K}} C_{l,k} \right).$$

Proposition 1. *The allocation $x(p)$ is a continuous function of λ , while $r(\lambda)$ and the aggregate excess demand $z(\lambda, \mathbf{t})$ are upper hemi-continuous correspondences of λ .*

Proof. See Appendix D.

Proposition 2. *For every $q \in C$, $x(\lambda'), r(\lambda')$ solves **Max 1**.*

Proof. The proof is the same in spirit as that of Proposition 4.3 in [23], therefore it is omitted.

We use these propositions to complete the proof of Lemma 2. We denote the point with label i of the primitive set W^D , by $q_{D,i}$. Since S is compact, for every $i \in \{0, 1, \dots, LK\}$, the sequence $\{q_{D,i}\}_D$ has a cluster point. As $D \rightarrow \infty$ the distance between the vertices of W^D goes to 0, so $\|q_{D,i} - q_{D,j}\| \rightarrow_D 0$ for any $i, j \in \{0, 1, \dots, LK\}$. This means that for any $i \in \{0, 1, \dots, LK\}$ the sequences $\{q_{D,i}\}_D$ have identical cluster points. Pick any such cluster point and denote it by λ . Since $z_{l,k}(\lambda, \mathbf{t})$ is an upper hemi-continuous correspondence (Proposition 1), λ is in C . This proves that C is non empty. The result of Lemma 2 follows from Proposition 2.

Proof Theorem 2. The assertion of the theorem is a direct consequence of Lemma 2.

In Section 5 we discuss cases where the convergence to the optimum value can be done in finite number of steps.

5 Discussion and conclusions

We presented an approach for optimal admission, resource allocation, and routing in multi-service connection-oriented networks. We described a con-

vergent and decentralized iterative procedure that leads to a solution of a fairly general admission, resource allocation and routing problem.

The main contributions of this paper are:

- (i) The generalization of the model of [23] and the solution of the routing problem within the context of the new model.
- (ii) The relaxation of assumptions on the constraint sets $\mathbf{F}^{i,j,t}$ that describe the allocations satisfying the QoS requirements for each type of service. Specifically, only the compactness of the sets $\mathbf{F}^{i,j,t}$ is required for the derivation of the results of this paper. The convexity of $\mathbf{F}^{i,j,t}$, which was needed in [23] and which is in general difficult to satisfy, is no further required.

The main features of our approach are the following:

- (1) The objective of the resource allocation process is to maximize the total value of the network to its users.
- (2) The users' utility functions u_i are quasi-linear, continuous, differentiable, locally non-satiated and strictly concave.
- (3) The agents are price takers in the markets in which they participate.
- (4) There is no cost associated with the supply of network resources.

An extensive discussion and critique of features 1, 3 and 4 has been presented in [23]. Thus, here we critique feature 2.

(2) Since, we assume that the expenditure of the good under study is a small portion of a consumer's total expenditure, the small size of the market under study should lead the prices of the other goods to be approximately unaffected by changes in this market. Because of this fixity of other prices, we are justified in treating the expenditure on these other goods as a single composite commodity, which we call the numeraire. This allows us to express the utility function as a function of the goods under study and the numeraire.

The choice of representing users' preferences by quasi-linear objective functions also imposes the constraint that there are no income effects on network service demand; that is, changes in income or budget available to the users does not change the amount of network services they wish to purchase. This is a typical simplifying assumption in the economic literature when the budget share of the services of interest is small, e.g. when network services are only a relatively small amount of the users' total expenditures.

The rest of the assumptions made for the utility functions are normal assumptions usually made in analysis of economic optimization problems. The continuous and differentiable assumption comes from the idea that we may look at a set of users that may have similar utilities as a group, and in this case the group utility will be a smoothed out version of each user's utility. Locally non-satiated and strictly concave assumptions are natural assumptions when we are working with goods that are desirable.

Next we comment on issues associated with our approach to the solution of the admission control, resource allocation and routing problem formulated in this paper. Specifically, we address the following:

- (a) What is the relation between a solution of **Max 1** and the choice of a particular **D**, defined in Section 4.3?
- (b) If during the Tatônnement process, described by Algorithm (**Q**), the minimum price of a particular service is achieved along more than one route, how is the route for that service chosen?

We discuss each issue separately.

(a) In this paper we proved the existence of a solution to the optimization problem **Max 1** and presented an algorithm (Algorithm (Q)) that converges to such a solution.

Algorithm (Q) works by taking a sequence of sets, denoted by $\{P^D\}_{D \in \mathbb{N}}$, of evenly distributed points in S . For each one of these sets the algorithm generates a primitive set W^D , satisfying the properties of Lemma 1. The theory tells us that as D goes to infinity, a subsequence W^D of primitive sets will contain a solution of **Max 1**. Since the size (diameter) of W^D goes to 0 as D goes to infinity, we are able to approximate a solution of **Max 1** by Algorithm (Q).

The question that remains is: “For a given D do we know if a solution of **Max 1** is contained in W^D , and if not, how far away is W^D from a solution of **Max 1**”? We can not answer this question in general. The answer depends on the behavior of the excess demand correspondence $z_{l,k}(\lambda, \mathbf{t})$, for all $l \in \mathbf{L}$ and $k \in \mathbf{K}$. Hence, without any further assumptions on the behavior of $z_{l,k}(\lambda, \mathbf{t})$, $l \in \mathbf{L}, k \in \mathbf{K}$ all we can guarantee is that in the limit, as $D \rightarrow \infty$, Algorithm (Q) will lead to a solution of **Max 1**. That is, as $D \rightarrow \infty$ the size of W^D decreases, and under certain regularity conditions (that relate the local maxima and local minima of $z_{l,k}(\lambda, \mathbf{T})$, $l \in \mathbf{L}, k \in \mathbf{K}$) W^D will contain a solution of **Max 1**.

In the case where the route for each service is fixed, as in [23], i.e. $\mathbf{T}^{i,j}$ is a singleton for all i, j , the excess demand correspondence will be a uniform continuous function (see [23]). If information on the local maxima and minima of $z_{l,k}(\lambda, \mathbf{t})$ is available, then the minimal D for which W^D contains a solution of **Max 1** can be calculated.

(b) If during the Tatônment process, described by Algorithm (Q), the minimum price of a particular service is achieved along more than one route, we have two options:

- (i) We can choose one of the aforementioned routes at random and proceed with the process.
- (ii) We can check the excess demand function $z_{l,k}(\lambda, \mathbf{t})$, $l \in \mathbf{L}, k \in \mathbf{K}$, corresponding to the choice of each combination of minimum priced routes. If for some choice we end up with an excess demand $z_{l,k}(\lambda, \mathbf{t}) \leq 0$, $\forall l \in \mathbf{L}, k \in \mathbf{K}$, then we have, for a very large D (see a), a good approximation to a solution of **Max 1**. Otherwise, we select a combination of routes of minimum price for each service and proceed with the Tatônment process.

According to the results of this paper, either option will eventually lead to a solution of **Max 1**. If we follow Option (ii) we may have to solve problems of the form (4.2) more often than we would have to with Option (i). On the other hand, with Option (i) we may have to solve problems of the form (4.4)–(4.5) more often than we would have to with Option (ii).

We conclude our discussion with two additional comments:

According to Algorithm (Q), after the auctioneer sets the prices for the resources, the service provider determines the minimal-cost route for each service and announces the corresponding prices per unit of service to the users. This feature of Algorithm (Q) is similar to that of the dual algorithm for the resource allocation and routing problem (without QoS requirements) presented in [10].

The class of networks to which the methodology presented in this paper applies includes those consisting of wired and wireless components.

A The auctioneer price adjustment mechanism

The price adjustment mechanism begins by fixing $D \in \mathbb{N}$, defining the $(LK + 1) \times (LK + 1)$ matrix

$$M_0 = \begin{bmatrix} D & D+1 & \cdots & D+1 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -1 \end{bmatrix} \quad (\text{A.1})$$

and letting

$$\lambda_{l,k}^{(0)} = M((l-1)K + k, 0) / M_0(0, 0), \quad l \in \mathbf{L}, k \in \mathbf{K},$$

where $M_0(i, j)$ denotes the $(i, j)^{\text{th}}$ entry of M_0 .

Starting with $\lambda_{l,k}^{(0)}$, $l \in \mathbf{L}, k \in \mathbf{K}$, the mechanism generates a sequence of prices $\lambda_{l,k}^{(m)}$, $l \in \mathbf{L}, k \in \mathbf{K}, m = 1, 2, \dots$, based on the construction of a sequence of $(LK + 1) \times (LK + 1)$ matrices $M_m, m = 1, 2, \dots$, where the original matrix is M_0 and each M_m is obtained from the previous one by replacing a single column. If c is the new column in M_m then

$$\lambda_{l,k}^{(m)} = M_m((l-1)K + k, c) / M_m(0, c), \quad l \in \mathbf{L}, k \in \mathbf{K}.$$

The replacement for the column c equals the sum of columns $c \ominus 1$ and $c \oplus 1$ minus column c itself, where the symbols \ominus, \oplus denote modulo LK subtraction and addition, respectively.

The rule determining the column that is replaced at each step consists of two stages: (1) a labelling procedure; and (2) a replacement process. We describe each stage separately:

1. Labelling procedure:

Initial Step: For $i \in \{0, 1, \dots, LK\}$, we label the i^{th} column of M_0 with the integer $v(i) \in \{0, 1, \dots, LK\}$. For $i = 1, \dots, LK$, we set $v(i) = i$. For $i = 0$ we observe the aggregate excess demand $z(\lambda)$ that results from the announcement of the price vector $\lambda_{l,k}^{(0)}$ to the market; if $z_{l,k}(\lambda) \leq 0$ for all $l \in \mathbf{L}, k \in \mathbf{K}$ then $v(0) = 0$; otherwise, $v(0) = \min\{(l-1)K + k : z_{l,k}(\lambda) > 0\}$.

Iteration Step: An integer $v(c) \in \{0, 1, \dots, LK\}$ is assigned to the new column c , that is brought into the matrix as a replacement, according to the following rule: If the column has a negative entry, then $v(c)$ is the smallest entry for which this is true. If the column has a zero entry, then $v(c)$ is the smallest entry for which this is true. If none of the above is true, then the label of column c is determined by the same procedure used for column 0 in the Initial Step, where the announced price vector is $\lambda_{l,k}^{(m)} = M((l-1)K + k, c) / M(0, c)$, $l \in \mathbf{L}, k \in \mathbf{K}$.

2. Replacement Process:

The matrix resulting after an integer, (label), $v(c)$ is assigned to a new column c , according to the labeling procedure, has one of the following two features:

- (F1) 1. None of the columns is associated with the label 0;
- 2. all of the columns have distinct labels, except for a single pair whose labels are identical;
- 3. one member of the pair of columns with identical labels has just been brought into the matrix.
- (F2) All columns of the matrix have distinct labels that span the set $\{0, 1, \dots, LK\}$ (in this case the column whose label is 0 has just been brought into the matrix).

If the matrix constructed has feature (F1), the column to be replaced is the one from the pair with identical labels that has *not* just been brought into the matrix. If the matrix constructed by the algorithm has feature (F2), no replacement is needed. Algorithm (Q) terminates.

B Algorithm (Q) from auctioneers' perspective

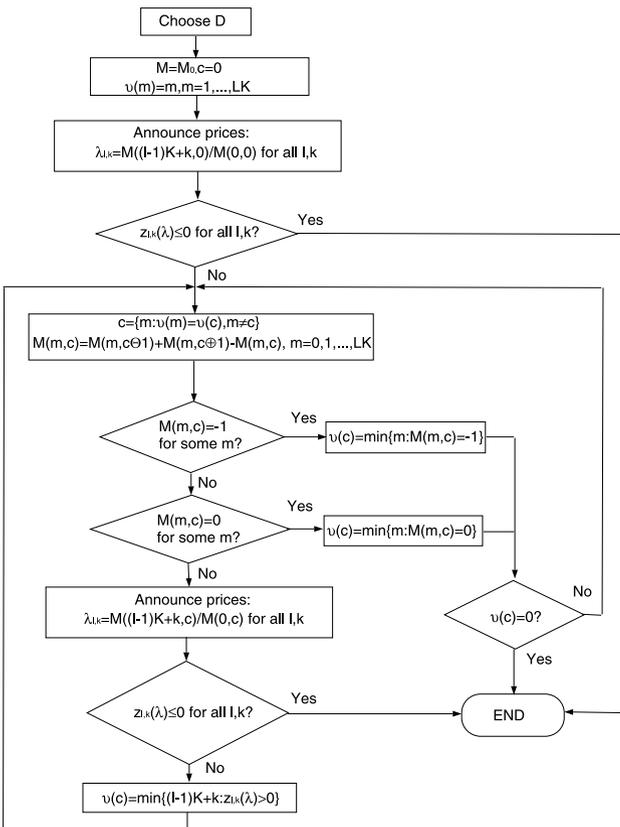


Fig. 1. Algorithm (Q) from the auctioneer's point of view

C Proof of Lemma 1

Proof. This proof is based on the following iterative idea:

(IP1)

1. Define a unique replacement operation for a single side or vector of a primitive set in S .
2. Show that for every vector in a primitive set the replacement algorithm uniquely determines the vector to be replaced with, and the new set of vectors forms a primitive set.
3. Start with the primitive set \mathcal{Q} . If all the vectors in \mathcal{Q} have distinct labels (i.e. if the interior vector has label 0) stop; otherwise, using the replacement operation in Step 1, replace the side that has the same label as the interior vector q .
4. If the new vector or side has a zero label, terminate the procedure. Otherwise, replace the element in the primitive set that has the same label with the element that was just brought in, and repeat Step 4.

We proceed to show that the above iterative process terminates after a finite number of steps and produces a primitive set in S^D such that its vectors have distinct labels. For any $q \in P^D$, q is of the form $(\frac{q_0}{D}, \dots, \frac{q_{LK}}{D})$, where q_1, \dots, q_{LK} , $D \in \mathbb{N}$. We let $\mathbb{T} : P^D \rightarrow \{\bar{z} : \bar{z} \in \mathcal{Z}_+^{LK+1}, \sum_{i=0}^{LK} z_i = D\}$, be the map defined by $\mathbb{T}(q) = (q_0, \dots, q_{LK})$. We let $\mathbb{T}(s^i) = \{\bar{z} : z_i \leq 0, \bar{z} \in \mathcal{Z}_+^{LK+1}, \sum_{i=0}^{LK} z_i = D\}$. Note that by the construction of T , any replacement algorithm of vectors in $\mathbb{T}(S^D)$, will induce via \mathbb{T}^{-1} a replacement algorithm of vectors in S^D .

We let the set \mathcal{M} be the set of $(KL + 1) \times (KL + 1)$ matrices with the column vectors being the \mathbb{T} image of vectors that form primitive sets in S^D . The form of these matrices is characterized in Theorem 6.2.1 through Theorem 6.2.9 [21].

The replacement algorithm is defined as follows. Start with $M \in \mathcal{M}$, and arrange all the columns vectors in increasing lexicographic order [21]. After reordering of the columns is done, the replacement algorithm replaces column i with the vector whose coefficients equal to the sum of column $i \ominus 1$ and $i \oplus 1$ minus the column i , where \ominus, \oplus are the (mod LK) subtraction and addition respectively.

Theorem 6.3.1 [21] proves that this replacement algorithm is unique, it is well defined (as long as we do not start with a primitive set of the form (s^i, \dots, s^{LK}, q) , with $s^i \in \mathbb{S}$ and $q \in P^D$, and we are trying to replace vector q), and that the set resulting after the replacement is a primitive set.

We now have defined the method of the replacement algorithm, and we proceed with the Steps 3–4 or the **(IP1)**. This procedure can only terminate if and only if we arrive at a desired primitive set (i.e. one with all distinct labels), or the replacement procedure can not be performed. In [21, pages 47–48] it is shown that we will never have the case that the replacement procedure is impossible, and also that the algorithm will never return to a primitive set that was previously visited. Since there are a finite number of possible primitive sets the algorithm will have to stop, so it will converge in finite number of steps to a primitive set with distinct labels.

D Proof of Proposition 1

Proof. We first proceed by proving that $r(\lambda)$ is an upper-hemicontinuous correspondence. Define the correspondence $\varphi^{i,j} : \mathbb{R}_+^{LK} \rightarrow \prod_{i=1}^m \mathbb{R}_+^{LK}$ where:

$$\begin{aligned} \varphi^{i,j}(\lambda) &= \{\bar{r}^{i,j} : r^{i,j,t} \in \mathbf{F}^{i,j,t}, 1 \leq t \leq |\mathbf{T}^{i,j}|\} \\ &= \prod_{t=1}^m \mathbf{F}^{i,j,t} \end{aligned} \tag{D.1}$$

Note that $\varphi^{i,j}$ is a constant relation, so it is continuous. Also $\varphi^{i,j}$ is compact valued since $\mathbf{F}^{i,j,t}$ is compact by assumption.

Given λ we define $f_\lambda : \prod_{i=1}^m \mathbb{R}_+^{LK} \rightarrow \mathbb{R}$ such that:

$$f_\lambda(\bar{r}^{i,j}) = - \min_{t \in \mathbf{T}^{i,j}} \left\{ \sum_{l \in \mathbf{L}} \sum_{k \in \mathbf{K}} \lambda_{l,k} r_{l,k}^{i,j,t} \right\}.$$

Define $r : \mathbb{R}_+^{LK} \rightarrow \prod_{i=1}^m \mathbb{R}_+^{LK}$ such that $r(\lambda) = \{y \in \varphi^{i,j}(\lambda) : y \text{ maximizes } f \text{ on } \varphi^{i,j}(\lambda)\}$. In other words $r(\lambda)$ is the set of sets of allocations that result in, for every service, a minimum price per unit of service.

The following theorem [1], known as Berge’s Maximum Theorem, shows that $r(\lambda)$ is upper hemicontinuous.

Theorem 3. *Let $\varphi : \mathbb{R}_+^{LK} \rightarrow \prod_{i=1}^m \mathbb{R}_+^{LK}$ be a compact-valued correspondence. Let $f : \prod_{i=1}^m \mathbb{R}_+^{LK} \rightarrow \mathbb{R}$ be continuous. Define $r : \mathbb{R}_+^{LK} \rightarrow \prod_{i=1}^m \mathbb{R}_+^{LK}$ by $r(\lambda) = \{y \in \varphi(\lambda) : y \text{ maximizes } f \text{ on } \varphi(\lambda)\}$, and $p : \mathbb{R}_+^{LK} \rightarrow \mathbb{R}$ by $p(\lambda) = f(y)$ for $y \in r(\lambda)$. If φ is continuous at λ , then r is closed and upper hemi-continuous at λ and p is continuous at λ . Furthermore, r is compact-valued.*

Next we show that $x(\lambda)$ is a continuous function of λ . Let $i \in \mathbf{N}$, and consider the following problem,

$$\max_{x^i \in X^i} \left[u_i(x^i) - \sum_{j \in \mathbf{M}_i} p_j^i x^{i,j} \right], \tag{D.2}$$

where, $p^{i,j} \triangleq \sum_{l \in \mathbf{L}} \sum_{k \in \mathbf{K}} \lambda_{l,k} r_{l,k}^{i,j}(\lambda)$.

The function $u_i(\cdot)$ is continuous and strictly convex on \mathbb{R}_+^M . Therefore, according to [16, Proposition 3.D.2(iii)] and [16, Proposition 3AA.1], the solution $x^i(p^i)$ to (D.2) is continuous at all vectors $p^i \triangleq \{p_j^i\}_{j \in \mathbf{M}_i}$, that satisfy $p_j^i > 0$, for all $j \in \mathbf{M}_i$. This result, along with the continuity of p^i with respect to λ , implies that $x(\lambda)$ is a continuous function of λ . The continuity of $x(\lambda)$ and the upper hemi-continuity of $r(\lambda)$, establish the fact that the aggregate excess demand function given by (4.10) is an upper hemi-continuous correspondence of λ . This completes the proof of Proposition 1.

Acknowledgement. This research was supported in part by NSF Grant ECS-9979347, and NSF Grant CCR-0082784.

References

1. Border KC (1989) Fixed point theorems with applications to economics and game theory. Cambridge University Press, New York
2. Cocchi R, Estrin D, Shenker S, Zhang L (1993) Pricing in computer networks: motivation, formulation, and example. *IEEE/ACM Transaction on Networking* 1(6):614–627, Dec
3. Courcoubetis C, Kelly FP, Weber RR (1997) Measurement-based usage charges in communications networks. Technical Report 19, Statistical Laboratory, University of Cambridge
4. de Veciana G, Baldick R (1998) Resource allocation in multiservice networks via pricing. *Computer Networks and ISDN Systems* 30:951–962
5. Gupta A, Stahl DO, Whinston AB (1997) A stochastic equilibrium model of internet pricing. *Journal of Economic Dynamics and Control* 21:697–672
6. Hurwicz L (1973) The design of mechanisms for resource allocation. *American Economic Review* 63(2):1–30
7. Jiang H, Jordan S (1995) The role of price in the connection establishment process. *European Transactions on Telecommunications* 6(4):421–429, July–Aug
8. Jordan S, Jiang H (1995) Connection establishment in high speed networks. *IEEE Selected Areas in Communications* 13(7):1150–1161, Nov
9. Kelly FP (1994) On tariffs, policing and admission control for multiservice networks. *Operations Research Letters* 15:1–9
10. Kelly FP, Maulloo AK, Tan DKH (1998) Rate control for communication networks: shadow prices, proportional fairness and stability. *Operational Research Society* 49:237–252
11. Kurose JF, Simha R (1989) A microeconomic approach to optimal resource allocation in distributed systems. *IEEE Transactions on Computers* 38(5):705–717
12. Low S, Varaiya P (1993) A new approach to service provisioning in ATM Networks. *IEEE/ACM Trans. Networking* 1:547–553, Oct. See also [13] for corrections
13. Low S, Varaiya P (1994) Corrections to: A new approach to service provisioning in ATM Networks. *IEEE/ACM Trans. Networking* 2:312, June
14. MacKie-Mason JK, Varian HR (1995) Pricing congestible network resources. *Journal of Selected Areas in Communications* 13(7):1141–1149, Sept
15. MacKie-Mason JK, Varian HR (1995) Some FAQs about usage-based pricing. *Computer Networks and ISDN Systems* 28:257–265
16. Mas-Colell A, Whinston MD, Green JR (1995) *Microeconomic theory*. Oxford University Press, New York
17. Murphy L, Murphy J (1994) Bandwidth allocation by pricing in ATM Networks. In *Proc. of the IFIP Broadband Communications*, March
18. Murphy L, Murphy J, Posner EC (1994) Distributed pricing for embedded ATM Networks. In *Proc. of the International Teletraffic Congress ITC-14*, June
19. Parris C, Ferrari D (1992) A resource based pricing policy for real-time channels in a packet-switching network. Technical report, International Computer Science Institute, Berkeley, CA
20. Parris C, Keshav S, Ferrari D (1992) A framework for the study of pricing in integrated networks. Technical Report TR-92-016, International Computer Science Institute, Berkeley, CA, March
21. Scarf H (1973) *The computation of economic equilibria*. Yale University Press, New Haven and London
22. Thomas P, Teneketzis D (1997) An approach to service provisioning with quality of service requirements in ATM Networks. *Journal of High Speed Networks* 6(4):263–291
23. Thomas P, Teneketzis D, MacKie-Mason JK (1999) A market – based approach to optimal resource allocation in integrated – services connection-oriented networks. Technical report, Systems Science and Engineering Division Department of Electrical Engineering, University of Michigan, <http://www-personal.umich.edu/jmm/research.html>
24. Viswanath P, Anantharam V, Tse DNC (1999) Optimal sequences, power control, and user capacity of synchronous cdma systems with linear mmse multiuser receivers. *IEEE Transactions on Information Theory* 45(6):1968–1983, September
25. Wang Q, Peha JM, Sirbu MA (1997) Optimal pricing for integrated-services networks. In: McKnight LW, Bailey JP (eds) *Internet Economics*, pages 353–376. MIT Press, 3 edition