Pricing in Multiservice Loss Networks: Static Pricing, Asymptotic Optimality, and Demand Substitution Effects

Ioannis Ch. Paschalidis and Yong Liu

IEEE/ACM Transactions On Networking, Vol. 10 (2002), No. 3, pp. 425–438.

Abstract— **We consider a communication network with fixed routing that can accommodate multiple service classes, differing in bandwidth requirements, demand pattern, call duration, and routing. The network charges a fee per call which can depend on the current congestion level, and which affects user's demand. Building on the single-node results of Paschalidis and Tsitsiklis, 2000, we consider both problems of revenue and welfare maximization and show that static pricing is asymptotically optimal in a regime of many, relatively small, users. In particular, the performance of an optimal (dynamic) pricing strategy is closely matched by a suitably chosen class-dependent static price, which does not depend on instantaneous congestion. This result holds even when we incorporate demand substitution effects into the demand model. More specifically, we model the situation where price increases for a class of service might lead users to use another class as an imperfect substitute. For both revenue and welfare maximization objectives we characterize the structure of the asymptotically optimal static prices, expressing them as a function of a parsimonious number of parameters. We employ a simulation-based approach to tune those parameters and to efficiently compute an effective policy away from the limiting regime. Our approach can handle large, realistic, instances of the problem.**

Keywords— **pricing, Internet economics, loss networks, revenue management, welfare maximization, Markov decision processes.**

I. Introduction

New application requirements and developments in Internet protocols are leading the way towards an "enhanced" next-generation Internet. This new medium will surpass the current "best effort only" capability and evolve into a *multiservice* network able to accommodate differentiated classes of service to support various types of applications and business requirements. In this new environment pricing of network services is becoming increasingly important: (i) it allows providers to recover their operating expenses and fund future capacity expansions, (ii) it can lead to more efficient use of the network resources by providing sufficient incentives to users, and (iii) enables the creation of a healthy market environment, where new network services can be (profitably) introduced and sustained.

In this paper, we consider a communication network with fixed routing that accommodates multiple service classes, differing in bandwidth requirements, demand pattern, call duration, and routing. Links in the network have given finite capacities and the total resource requirement of all calls using a link cannot exceed the link's capacity. The network charges a fee per call which can depend on the current congestion level, and which affects user's demand for calls. We propose pricing strategies that aim at two distinct objectives: either maximizing the revenue of the network operator or maximizing the social welfare of users.

Pricing in communication networks has received a lot of attention in the literature. MacKie-Mason and Varian [1] proposed a "smart market" where individual packets bid for transport while the network only serves packets with bids above a certain (congestion-dependent) cutoff amount. Kelly et al. [2], [3] consider charges that increase with either realized flow rate or with the "share" of the network consumed by a traffic stream. Several researchers have looked at packet-based pricing schemes as an incentive for more efficient flow control (see e.g., Gibbens and Kelly [4], La and Anantharam [5], Kunniyur and Srikant [6]). Equilibrium properties of bandwidth and buffer allocation schemes are analyzed by Low [7].

The network model we consider in this paper is more appropriate for real-time traffic that requires strict *Quality of Service (QoS)* guarantees. Such guarantees can often be translated into a preset resource amount that has to be allocated to a call at all links in its route through the network. If the resource is bandwidth this resource amount can be some sort of an *effective bandwidth* (see e.g., Kelly [8] for a survey of effective bandwidth characterizations and Paschalidis [9] for similar notions in the multiclass case). In this setting, Kelly [10] and Courcoubetis et al. [11] propose the pricing of real-time traffic with QoS requirements, in terms of its effective bandwidth.

The revenue maximization perspective we take is influenced by similar work in the revenue management of airline reservations (see, e.g., Gallego and van Ryzin [12]). Technically, the problem we consider is different; it considers a long-term average vs. a finite horizon setup.

Our work is closer to Paschalidis and Tsitsiklis [13] that considered pricing of multiple services sharing a single resource. In fact, we generalize the main result of [13] in

Revised August 6, 2002. This is a slightly revised version of the paper that appeared in print.

I. Ch. Paschalidis is with the Center for Information and Systems Engineering (CISE) and the Department of Manufacturing Engineering, Boston University, Boston, MA 02215, e-mail: yannisp@bu.edu, url: http://ionia.bu.edu/. Research partially supported by the NSF under a CAREER award ANI-9983221 and grants NCR-9706148 and ACI-9873339, by the ARO under the ODDR&E MURI2001 Program Grant DAAD19-01-1-0465 to the Center for Networked Communicating Control Systems, and by the Nokia Research Center (Boston).

Yong Liu is with the Department of Manufacturing Engineering, Boston University, Boston, MA 02215, e-mail: liuyong@bu.edu.

several directions:

- *A Network Setting.* The network model we consider is what Kelly [14] calls a *loss network* with fixed routing (see also Ross [15]). We show that in a limiting regime of "many small users," laws of large numbers take effect and a simple *static* pricing scheme is asymptotically optimal. That is, under stationarity assumptions, prices can be class-dependent but remain fixed; it is not necessary to employ a dynamic scheme according to which prices depend on the congestion level. If demand is nonstationary and characterized by time-of-day effects, which is widely agreed to be the case, the proposed pricing scheme leads to *time-of-day pricing.* The "many small users" regime we consider is appropriate for large networks such as the Internet where (backbone) capacities are large and individual sessions occupy a small fraction of those capacities.
- *Rate of convergence.* We characterize the rate at which a static pricing policy converges to the optimal in the regime of many small users. This allows us to obtain bounds on the suboptimality gap of static pricing away from the limiting regime. We provide examples where such bounds are useful in quickly assessing efficiency gains achieved by appropriately scaling the system.
- *Demand substitution effects.* We extend the basic demand model of [13] to cover the case where users might decide to use another class as an imperfect substitute, when they perceive their desired class to be expensive. For this model as well, we show that static pricing is asymptotically optimal in the regime of many small users. To that end, we rely on asymptotic results for blocking probabilities from [14]. Our modeling of demand substitution effects is in fact similar to the single-link model of Courcoubetis and Reiman [16]; our work can be seen as a generalizing theirs in a network setting.

A static pricing scheme, such as the one we propose, has obvious implementation advantages: charges are predictable by users, evolve in a slower time-scale than congestion phenomena, and no elaborate real-time mechanism is needed to communicate prices to the users. Moreover, as we will see, prices can be computed in large scale systems, which is not the case with the optimal dynamic pricing scheme. To that end, from our asymptotic optimality results we first identify an insightful, asymptotically optimal, structure of static prices under both revenue and welfare maximization objectives. According to this structure prices depend on a parsimonious number of parameters. We then employ a simulation-based optimization technique to tune those parameters. We report results from a number of numerical experiments, including, a large scale one, indicating that this approach yields near-optimal policies.

The network model we propose is general enough to accommodate several situations of practical interest. It can be seen as modeling the pricing of bandwidth by a network provider who offers a menu of services to users. Users can in fact also be smaller "retail" providers, in which case calls can be seen as virtual circuits leased from the backbone provider. The model can also be seen as pricing the use of Web or other servers by an application service provider: a "call" is associated with a transaction that requires cooperation from a series of servers, thus, it ties up a fraction of their capacities until it is completed.

The remainder of this paper is organized as follows. In Section II we introduce the basic model and formulate both problems of revenue and welfare maximization. We consider the optimal dynamic policy and some of its properties in Section III. A static pricing policy is introduced in Section IV. In Section VI we prove the asymptotic optimality of static prices in the regime of many small users. The proof is based on an upper bound on optimal performance we develop in Section V and allows us to obtain performance guarantees on the suboptimality gap of static pricing policies. Section VII contains the treatment of a more general demand model that incorporates demand substitution effects. Section VIII discusses the computation of static policies in large scale networks. Numerical results are presented in Section IX and conclusions are in Section X.

II. The Network Model

In this section we will introduce the model of the multiservice network we wish to study. We consider a network with L links. The capacity of each link j is C_j units of bandwidth for $j = 1, \ldots, L$. We will write $\mathbf{C} = (C_1, \ldots, C_L)$. On a notational remark, we will be denoting all vectors using boldface and assume that they are column vectors unless otherwise explicitly specified. For economy of space, we will be writing $\mathbf{x} = (x_1, \ldots, x_m)$ to identify the elements of a vector $\mathbf{x} \in \mathbb{R}^m$. The network provides M classes of service. Each service class is distinguished by its demand pattern, bandwidth requirement, call duration, and routing through the network. Classes have a fixed route through the network. In particular, class i requires r_{ji} units of bandwidth from link j, for $i = 1, \ldots, M$ and $j = 1, \ldots, L$. The routing matrix will be denoted by $\mathbf{R} = \{r_{ji}\}$, i.e., an $L \times M$ matrix with the (j, i) element being equal to r_{ji}. The route of class i is characterized by the sequence of links $j_{i_1}, j_{i_2}, \ldots, j_{i_l}$ it traverses; we will denote it by

$$\mathcal{R}_i = \big\{(j_{i_1}, j_{i_2}, \ldots, j_{i_l}) \mid 1 \leq j_{i_1}, \ldots, j_{i_l} \leq L,\\ r_{ji_k} > 0, k = 1, \ldots, l\big\}, \quad i = 1, \ldots, M.$$

We will write $j \in \mathcal{R}_i$ if link j is any link in the sequence $(j_{i_1}, j_{i_2}, \ldots, j_{i_l})$. To exclude trivial cases, we will be assuming that $\mathcal{R}_i \neq \emptyset$ for each class i. For all other links j that are not in route $\mathcal{R}_i$ it is understood that $r_{ji} = 0$.

We assume that calls of class $i = 1, \ldots, M$ arrive according to a Poisson process and stay in the system for a time interval which is exponentially distributed with rate μ_i. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$. The network charges a fee u_i per call of class i, which can depend on the current congestion level and which affects user's demand for calls. We will assume that the arrival rate of class i calls is a known function of u_i, which will be referred to as *demand function* and denoted by $\lambda_i(u_i)$. We will write $\mathbf{u} = (u_1, \ldots, u_M)$ and $\boldsymbol{\lambda}(\mathbf{u}) = (\lambda_1(u_1), \ldots, \lambda_M(u_M))$. We will be making the following assumption for demand functions.

Assumption A
For every $i = 1, \ldots, M$, $\lambda_i(u_i) \geq 0$, and there exists a price $u_{i,\max}$ beyond which $\lambda_i(u_i)$ becomes zero. Furthermore, the function $\lambda_i(u_i)$ is continuous and strictly decreasing in the range $u_i \in [0, u_{i,\max}]$.

Hence, the demand is at its peak when prices are zero. We will use $\boldsymbol{\lambda}_0 = (\lambda_{0,1}, \ldots, \lambda_{0,M}) \triangleq \boldsymbol{\lambda}(\mathbf{0})$ to denote the peak demand vector, where $\mathbf{0}$ is the vector of all zeroes.

Let $n_i(t)$ be the number of class i calls that are in progress at time t. We will make the convention that $n_i(t)$ is a right-continuous function of time. We will denote by $\mathbf{n}(t) = (n_1(t), \ldots, n_M(t))$ the state of the system at time t. An incoming class i call is accepted if all the links along its route have enough available bandwidth, that is, if $\mathbf{R}(\mathbf{n}(t) + \mathbf{e}_i) \leq \mathbf{C}$, where $\mathbf{e}_i$ is the ith unit vector, namely, a vector with all its components zero except the ith component which is equal to one. If this latter condition is violated, an incoming call is rejected and lost for the system. Let $\mathcal{S} = \{\mathbf{n} \mid \mathbf{Rn} \leq \mathbf{C}\}$ denote the state space for the system, i.e., the set of states at which capacity constraints are satisfied.

A *pricing policy* is a rule that determines the pricing vector $\mathbf{u}(t) = (u_1(t), \ldots, u_M(t))$ at any time t as a function of the state $\mathbf{n}(t)$. Under the assumptions put in place, for any given pricing policy the system evolves as a continuous-time Markov chain with state $\mathbf{n}(t) \in \mathcal{S}$. As in [13] we are interested in pricing policies for two distinct objectives: revenue maximization and social welfare maximization.

A. The Revenue Maximization Problem

Let us fix a pricing policy $\mathbf{u}(t)$. Assuming that there is enough bandwidth to accept a class i call, the instantaneous expected revenue rate from those calls is $\lambda_i(u_i(t))u_i(t)$, since class i arrivals are Poisson with rate $\lambda_i(u_i(t))$. If there is not sufficient bandwidth to accept class i calls we can, without loss of generality, set $u_i(t) = u_{i,\max}$ and bring the instantaneous expected revenue rate to zero. Thus, the total expected long-term average revenue is

$$\begin{aligned} J &= \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{M} \mathbf{E}\left[\int_0^T \lambda_i(u_i(t))u_i(t)dt\right] \\ &= \lim_{T\to\infty} \frac{1}{T} \mathbf{E}\left[\int_0^T \boldsymbol{\lambda}(\mathbf{u}(t))'\mathbf{u}(t)dt\right]. \end{aligned} \tag{1}$$

B. The Welfare Maximization Problem

To formulate the welfare maximization problem, we will interpret the demand model as follows. Potential calls of class i are generated according to a Poisson process with constant rate $\lambda_{0,i}$, which is the peak arrival rate of class i introduced earlier. A potential class i call, if it goes through, results in a user utility of U_i, where U_i is a non-negative random variable taking values in $[0, u_{i,\max}]$. Let $f_i(u_i)$ be the continuous probability density function of U_i. We assume that a potential class i call decides to join the system if and only if the utility it will extract, U_i, exceeds the prevailing price u_i. This implies that class i calls are realized according to a randomly modulated Poisson process with rate $\lambda_i(u_i(t)) = \lambda_{0,i}\mathbf{P}[U_i \geq u_i(t)]$. Furthermore, the expected utility, conditioned on the fact that a call has been established, is equal to $\mathbf{E}[U_i \mid U_i \geq u_i]$ under a current price of u_i. Hence, the expected long-term average rate at which utility is generated is given by

$$\lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{M} \mathbf{E}\left[\int_0^T \lambda_i(u_i(t))\mathbf{E}[U_i \mid U_i \geq u_i(t)]dt\right]. \tag{2}$$

This is an objective of exactly the same form as in the case of revenue maximization, except that the instantaneous revenue rate $\lambda_i(u_i)u_i$ of class i is replaced by $\lambda_i(u_i)\mathbf{E}[U_i \mid U_i \geq u_i(t)]$. Thus, the two problems can be treated using the same set of tools. According to the utility assumptions put in place we have:

$$\lambda_i(u_i) = \lambda_{0,i} \int_{u_i}^{u_{i,\max}} f_i(v)dv, \tag{3}$$

and

$$\lambda_i(u_i)\mathbf{E}[U_i \mid U_i \geq u_i] = \lambda_{0,i} \int_{u_i}^{u_{i,\max}} v f_i(v)dv. \tag{4}$$

III. Optimal Dynamic Policy

We will start the analysis by considering optimal (dynamic) pricing policies. Under both objectives of revenue and welfare maximization the problem can be formulated using stochastic dynamic programming (DP). We first consider revenue maximization.

The state of the system $\mathbf{n}(t)$ evolves as a continuous-time Markov chain and its total transition rate out of any state is bounded by

$$\nu = \sum_{i=1}^{M} \left(\lambda_{0,i} + \mu_i \max_{j\in\mathcal{R}_i} \left\lceil \frac{C_j}{r_{ji}} \right\rceil\right).$$

The Markov chain can be uniformized, leading to a Bellman equation of the form

$$\begin{aligned} J^* + h(\mathbf{n}) = \max_{\mathbf{u}\in\mathcal{U}} \Bigg[&\sum_{i\notin\mathcal{C}(\mathbf{n})} \lambda_i(u_i)u_i \\ &+ \sum_{i\notin\mathcal{C}(\mathbf{n})} \frac{\lambda_i(u_i)}{\nu} h(\mathbf{n} + \mathbf{e}_i) + \sum_{i=1}^{M} \frac{n_i\mu_i}{\nu} h(\mathbf{n} - \mathbf{e}_i) \\ &+ \left(1 - \sum_{i\notin\mathcal{C}(\mathbf{n})} \frac{\lambda_i(u_i)}{\nu} - \sum_{i=1}^{M} \frac{n_i\mu_i}{\nu}\right) h(\mathbf{n}) \Bigg], \end{aligned} \tag{5}$$

where $\mathcal{U} = \{\mathbf{u} \mid 0 \leq u_i \leq u_{i,\max}, \forall i\}$ is the set of possible price vectors and $\mathcal{C}(\mathbf{n}) = \{i \mid \mathbf{R}(\mathbf{n} + \mathbf{e}_i) \not\leq \mathbf{C}\}$ is the set of classes whose calls cannot be admitted in state $\mathbf{n}$. Here, J^* and $h(\mathbf{n})$ denote the optimal expected revenue rate and the so called relative reward in state $\mathbf{n}$ (see Bertsekas [17]). This DP formulation is in fact almost identical to the one in [13], the only difference being the definition of $\mathcal{C}(\mathbf{n})$ which has been extended to the network setting. It has been argued there that the standard infinite-horizon average-cost dynamic programming theory applies (see [17]), thus, there exists a stationary policy which is optimal. We will use $\mathbf{u}^*(\mathbf{n})$ to denote an optimal policy to explicitly indicate its dependence on the state of the system. Such a policy can be found by solving Bellman's equation using standard DP algorithms. However, Bellman's "curse of dimensionality" prohibits us from solving realistic instances of the problem. Consequently, we are interested in exploring simpler, yet not too far from the optimal, alternatives. Before we proceed with this agenda we state some properties of the optimal policy. These properties are simple extensions of the results in [13] for the single-link system, thus, we omit the proofs.

Theorem 1 *1.* **(Monotonicity of $h(\mathbf{n})$)** *For all i and all $\mathbf{n}$ such that $\mathbf{R}(\mathbf{n}+\mathbf{e}_i) \leq \mathbf{C}$, we have $h(\mathbf{n}) \geq h(\mathbf{n}+\mathbf{e}_i)$, where $\mathbf{e}_i$ denotes the ith unit vector.*
2. **(The infinite bandwidth case)** *If there are no capacity constraints on all links in the network (i.e., $C_j = \infty$, $\forall j$), the optimal revenue is given by*

$$J_\infty = \max_{\mathbf{u}\in\mathcal{U}} \sum_{i=1}^{M} \lambda_i(u_i)u_i,$$

and the optimal price vector is some constant $\mathbf{u}_\infty$ that does not depend on the state $\mathbf{n}$. Furthermore, we have $J^* \leq J_\infty$.
3. There exists an optimal policy $\mathbf{u}^*$ such that for every state $\mathbf{n}$, we have $\mathbf{u}^*(\mathbf{n}) \geq \mathbf{u}_\infty$.

The case of welfare maximization, can be treated similarly. Bellman's equation remains the same, except that the reward rate $\boldsymbol{\lambda}(\mathbf{u})'\mathbf{u}$ is replaced by $\sum_i \lambda_i(u_i)\mathbf{E}[U_i \mid U_i \geq u_i]$. As in Theorem 1 (1), the relative rewards $h(\mathbf{n})$ are again monotonically non-increasing in $\mathbf{n}$. If the bandwidth is infinite, welfare is maximized by admitting every user, and the optimal price $\mathbf{u}_\infty$ is equal to zero. For a finite capacity network, the optimal prices are non-negative, which provides a trivial extension of Theorem 1 (3).

IV. A Static Pricing Policy

Possibly the simplest pricing policy is a *static* policy, defined as the policy under which prices are fixed to some vector $\mathbf{u}$ independent of the state of the system. According to this policy the system evolves as a continuous-time Markov chain which has a unique stationary distribution. In particular, the steady-state distribution has a product form and under a static pricing policy $\mathbf{u}$ is given by (see Kelly [14] and Ross [15])

$$\pi_{\mathbf{n}}(\mathbf{u}) = \mathbf{P}[\mathbf{n}(t) = \mathbf{n} \mid \mathbf{u}(t) = \mathbf{u}] = \frac{1}{G(\mathbf{u})} \prod_{i=1}^{M} \frac{(\rho_i(u_i))^{n_i}}{n_i!}, \quad \mathbf{n} \in \mathcal{S}, \quad (6)$$

where $G(\mathbf{u})$ is a normalizing constant given by

$$G(\mathbf{u}) = \sum_{\mathbf{n}\in\mathcal{S}} \prod_{i=1}^{M} \frac{(\rho_i(u_i))^{n_i}}{n_i!},$$

and $\rho_i(u_i) = \lambda_i(u_i)/\mu_i$ is the load offered by class i.

According to the static pricing policy the prices stay fixed which results in a constant arrival rate $\boldsymbol{\lambda}(\mathbf{u})$ independent of the state of the system. As a result we can not eliminate demand by raising prices when available resources are not sufficient to accept a call. Thus, deviating from our earlier convention, we will be blocking calls that arrive to find no sufficient resources. Consequently, the blocking probability has to be taken into account when calculating revenue. The blocking probability of class i calls under the static policy $\mathbf{u}$ is given by

$$\mathbf{P}^i_{\text{loss}}(\mathbf{u}) = \sum_{\{\mathbf{n} \mid \mathbf{R}(\mathbf{n}+\mathbf{e}_i)\not\leq\mathbf{C}\}} \pi_{\mathbf{n}}(\mathbf{u}). \quad (7)$$

The optimal revenue by a static policy is given by

$$J_s = \max_{\mathbf{u}\in\mathcal{U}} J(\mathbf{u}) = \max_{\mathbf{u}\in\mathcal{U}} \sum_{i=1}^{M} \lambda_i(u_i)u_i \left(1 - \mathbf{P}^i_{\text{loss}}(\mathbf{u})\right), \quad (8)$$

and it can be no better than the optimal (dynamic) revenue, i.e., $J_s \leq J^*$.

The calculation of the optimal static revenue J_s and the corresponding optimal static policy $\mathbf{u}_s$ suffers from a similar "curse of dimensionality" as in the case of dynamic policies. In particular, to calculate the blocking probability one needs to compute the steady-state probabilities $\pi_{\mathbf{n}}(\mathbf{u})$ which depend on the normalizing constant G. Computing this constant for networks with arbitrary topologies is an NP-complete problem (see Louth [18]). Efficient schemes exist for special topologies and the so call *reduced load approximation* can be used to approximate the blocking probabilities in arbitrary networks [15]. Numerical difficulties, though, exist for the reduced load approximation in large systems. To overcome high dimensionality problems we are interested in scalable and efficient ways of computing "good" static policies.

For the case of welfare maximization, the same discussion applies, with $\lambda_i(u_i)u_i$ replaced by $\lambda_i(u_i)\mathbf{E}[U_i \mid U_i \geq u_i]$.

V. An Upper Bound on the Optimal Performance

We will next develop an upper bound on the optimal revenue J^*. Such a bound is useful because it can help us bound the suboptimality gap of suboptimal policies we consider in this paper. It will also be instrumental in establishing our asymptotic optimality results.

Let us denote by $u_i(\lambda_i)$ the inverse of the demand function $\lambda_i(u_i)$, which exists due to Assumption A. Let us also define $F_i(\lambda_i) \triangleq \lambda_i u_i(\lambda_i)$ and $F_i(\lambda_i) \triangleq \lambda_i \mathbf{E}[U_i \mid U_i \geq u_i(\lambda_i)]$, $i = 1,\ldots,M$, for the case of revenue and welfare maximization, respectively. We assume that the functions F_i are concave. This is true, for example, when the demand function $\lambda_i(u_i)$ is linear. The following theorem provides an upper bound on J^*.

Theorem 2 *Consider the following nonlinear optimiza-*

tion problem

$$\begin{aligned} &\textit{maximize} && \sum_{i=1}^{M} F_i(\lambda_i) \\ &\textit{subject to} && \lambda_i = n_i \mu_i, \qquad i = 1, \ldots, M, \\ & && \sum_i n_i r_{ji} \le C_j, \qquad j = 1, \ldots, L, \end{aligned} \tag{9}$$

and let $J_{\rm ub}$ *denote the optimal objective value. If* $F_i(\lambda_i)$ *is a concave function for all* $i = 1, \ldots, M$*, then* $J^* \le J_{\rm ub}$*.*

Proof: Consider an optimal dynamic pricing policy $\mathbf{u}^*$. Without loss of generality, we assume that the price u_i^* becomes large enough (e.g., $u_{i,\max}$) and the arrival rate $\lambda_i(u_i^*)$ is equal to zero, whenever the state $\mathbf{n}$ is such that a class i call cannot be admitted. In the system operating under the optimal policy, we can view the arrival rate, λ_i, and the number of class i customers in the system, n_i, as random variables. Let $\mathbf{E}[\cdot]$ denote the expectation with respect to the steady-state distribution under this particular policy $\mathbf{u}^*$. At any time, we have $\sum_i n_i r_{ji} \le C_j$, $\forall j$, which implies that $\sum_i \mathbf{E}[n_i] r_{ji} \le C_j$, $\forall j$. Furthermore, Little's law implies $\mathbf{E}[\lambda_i] = \mu_i \mathbf{E}[n_i]$. Thus, $\mathbf{E}[n_i]$, $\mathbf{E}[\lambda_i]$, $i = 1, \ldots, M$, form a feasible solution of the problem in (9). Using the concavity of F_i and Jensen's inequality, we have

$$J_{\rm ub} \ge \sum_{i=1}^{M} F_i(\mathbf{E}[\lambda_i]) \ge \sum_{i=1}^{M} \mathbf{E}[F_i(\lambda_i)] = J^*,$$

where the last equality used the optimality of the policy under consideration. ■

VI. Asymptotic Optimality of Static Pricing

We will now proceed with establishing our main results for the model considered in Section II, namely, the asymptotic optimality of static pricing and the derivation of guarantees on the suboptimality gap away from the limiting regime.

The limiting regime we will consider is one of *"many small users,"* in the sense that link capacities become large compared to the bandwidth of a typical call. More specifically, we start with a base system with finite demand function $\boldsymbol{\lambda}(\mathbf{u})$ and finite capacity $\mathbf{C}$ and then scale by increasing both demand and capacity by a scaling factor $c \ge 1$. We will use a superscript c to denote various quantities in the scaled system. In particular, in the scaled system the capacity is $\mathbf{C}^c = c(C_1, \ldots, C_L)$ and the demand function is given by $\boldsymbol{\lambda}^c(\mathbf{u}) = c(\lambda_1(u_1), \ldots, \lambda_M(u_M))$. Note that in the revenue maximization problem we simply scale the given demand function. In the welfare maximization problem it suffices to scale the peak demand rate as $\boldsymbol{\lambda}_0^c = c\boldsymbol{\lambda}_0$ and keep unaltered the behaviour of the users summarized in the utility density function $f_i(u_i)$ (see Section II-B). This results in a demand function $\lambda_i^c(u_i) = c\lambda_{0,i}\mathbf{P}[U_i \ge u_i]$. The remaining system parameters $\boldsymbol{\mu}$ and $\mathbf{R}$ are held fixed. The base system corresponds to the case $c = 1$.

In the scaled system the upper bound, $J_{\rm ub}^c$, is obtained by maximizing $\sum_i c\lambda_i(u_i)u_i$ in the revenue maximization case and $\sum_i c\lambda_i(u_i)\mathbf{E}[U_i \mid U_i \ge u_i]$ in the welfare maximization case. The constraints in the upper bound calculation become

$$\sum_i \frac{c\lambda_i(u_i) r_{ji}}{\mu_i} \le cC_j, \qquad \forall j, \tag{10}$$

which are identical to the constraints for the base system (cf. (9)). Hence, there exists an optimal solution $\mathbf{u}_{\rm ub}^* = (u_{{\rm ub},1}^*, \ldots, u_{{\rm ub},M}^*)$, which is independent of c, and it holds that $J_{\rm ub}^c = cJ_{\rm ub}^1$. In proving our asymptotic optimality result we will first consider the blocking probabilities in the scaled system. We will use the convention that for any static policy $\mathbf{u}$ for which $\lambda_i(u_i) = 0$, $\mathbf{P}_{\rm loss}^i(\mathbf{u}) = 0$. We will denote by $\mathcal{O}$ the set of classes with nonzero demand at $\mathbf{u}_{\rm ub}^*$, i.e., $\mathcal{O} = \{i \in \{1, \ldots, M\} \mid \lambda_i(u_{{\rm ub},i}^*) > 0\}$. We will also denote by $\mathcal{O}_j$, $j = 1, \ldots, L$, the set of classes $i \in \mathcal{O}$ that use link j, i.e., $\mathcal{O}_j = \{i \in \mathcal{O} \mid r_{ji} > 0\}$. We will assume that $\mathcal{O} \ne \emptyset$; otherwise $J_{\rm ub} = 0$ which can only happen in the trivial cases that $\mathbf{C} = \mathbf{0}$ or $\lambda(u_i) = 0$ for all u_i and i. Recall also that we have assumed $\mathcal{R}_k \ne \emptyset$; otherwise class k can be eliminated from the system. All classes $i \notin \mathcal{O}$ are shut out of the system under $\mathbf{u}_{\rm ub}^*$, do not contribute to the revenue or the social welfare, and according to our convention have zero blocking probability.

Proposition 1 *Consider either the revenue maximization problem or the welfare maximization problem and let* $\mathbf{u}_{\rm ub}^*$ *be the optimal solution to the upper bound problem in the scaled system with parameter* c*. For any* $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_M) > \mathbf{0}$*, consider the static policy* $\mathbf{u}^{\boldsymbol{\varepsilon}}$ *given by* $u_i^{\boldsymbol{\varepsilon}} = u_{{\rm ub},i}^* + \varepsilon_i$*,* $i = 1, \ldots, M$*. Let* $\mathbf{P}_{\rm loss}^{k,c}(\mathbf{u}^{\boldsymbol{\varepsilon}})$ *be the blocking probability of class* k *calls in the scaled system, under policy* $\mathbf{u}^{\boldsymbol{\varepsilon}}$*. For every class* $k \in \mathcal{O}$ *and all* c*, we have*

$$\mathbf{P}_{\rm loss}^{k,c}(\mathbf{u}^{\boldsymbol{\varepsilon}}) \le \sum_{j \in \mathcal{R}_k} \exp\left\{ \inf_{\theta \ge 0} \xi_{jk}^{\boldsymbol{\varepsilon}}(c, \theta) \right\}, \tag{11}$$

where

$$\xi_{jk}^{\boldsymbol{\varepsilon}}(c, \theta) \triangleq c \sum_{i \in \mathcal{O}_j} \Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta) + \theta r_{jk}, \tag{12}$$

and

$$\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta) \triangleq \frac{\lambda_i(u_i^{\boldsymbol{\varepsilon}})(e^{\theta r_{ji}} - 1) - \theta r_{ji} \lambda_i(u_{{\rm ub},i}^*)}{\mu_i}. \tag{13}$$

Furthermore, for all $k \in \mathcal{O}$ *and* $j \in \mathcal{R}_k$*,* $\inf_{\theta \ge 0} \xi_{jk}^{\boldsymbol{\varepsilon}}(c, \theta) \to -\infty$ *as* $c \to \infty$ *and*

$$\lim_{c \to \infty} \mathbf{P}_{\rm loss}^{k,c}(\mathbf{u}^{\boldsymbol{\varepsilon}}) = 0. \tag{14}$$

Proof: Since $\boldsymbol{\varepsilon} > \mathbf{0}$ and due to Assumption A, $\lambda_i^c(u_i^{\boldsymbol{\varepsilon}}) = 0$ for all $i \notin \mathcal{O}$ and all c. Thus, no customer exists in the system from those classes $i \notin \mathcal{O}$ and according to our convention the corresponding blocking probabilities are zero. We will next concentrate on classes $i \in \mathcal{O}$.

Let n_i^c (respectively $n_{i,\infty}^c$) be the random variable which is equal to the number of active class i calls, in steady-state, in the scaled system, under prices $\mathbf{u}^{\boldsymbol{\varepsilon}}$ and with capacity $c\mathbf{C}$ (respectively, with infinite capacity). By defining the arrival processes in these two systems on a common probability space we can see that for all sample paths n_i^c is smaller than $n_{i,\infty}^c$. Using this fact, for any class $k \in \mathcal{O}$ we have

$$\begin{aligned} \mathbf{P}_{\text{loss}}^{k,c}(\mathbf{u}^{\boldsymbol{\varepsilon}}) &= \mathbf{P}\left[\bigcup_{j\in\mathcal{R}_k}\sum_{i\in\mathcal{O}_j} r_{ji}n_i^c > cC_j - r_{jk}\right] \qquad (15)\\ &\leq \mathbf{P}\left[\bigcup_{j\in\mathcal{R}_k}\sum_{i\in\mathcal{O}_j} r_{ji}n_{i,\infty}^c > cC_j - r_{jk}\right]. \end{aligned}$$

In the above, note that since $k \in \mathcal{O}$ and $\mathcal{R}_k \neq \emptyset$, there exists at least one $j \in \mathcal{R}_k$ and $\mathcal{O}_j$ contains at least class k. Using the fact that $\mathbf{u}_{\text{ub}}^*$ satisfies the constraint (10), we obtain

$$\begin{aligned} &\mathbf{P}\left[\bigcup_{j\in\mathcal{R}_k}\sum_{i\in\mathcal{O}_j} r_{ji}n_{i,\infty}^c > cC_j - r_{jk}\right] \qquad (16)\\ &\leq \mathbf{P}\left[\bigcup_{j\in\mathcal{R}_k}\sum_{i\in\mathcal{O}_j} r_{ji}n_{i,\infty}^c > \sum_{i\in\mathcal{O}_j}\frac{c\lambda_i(u_{\text{ub},i}^*)r_{ji}}{\mu_i} - r_{jk}\right]\\ &\leq \sum_{j\in\mathcal{R}_k}\mathbf{P}\left[\sum_{i\in\mathcal{O}_j} r_{ji}n_{i,\infty}^c > \sum_{i\in\mathcal{O}_j}\frac{c\lambda_i(u_{\text{ub},i}^*)r_{ji}}{\mu_i} - r_{jk}\right], \end{aligned}$$

where the last inequality is due to the union bound.

Note next that the random variable $n_{i,\infty}^c$ is equal to the number of customers in an $M/M/\infty$ queue with arrival rate $c\lambda_i(u_i^{\boldsymbol{\varepsilon}})$ and service rate μ_i for each server. Its moment-generating function is

$$\mathbf{E}[e^{\theta n_{i,\infty}^c}] = e^{\frac{c\lambda_i(u_i^{\boldsymbol{\varepsilon}})}{\mu_i}(e^{\theta}-1)},$$

and by independence we obtain

$$\mathbf{E}\left[e^{\sum_{i\in\mathcal{O}_j}\theta r_{ji}n_{i,\infty}^c}\right] = \exp\left\{c\sum_{i\in\mathcal{O}_j}\frac{\lambda_i(u_i^{\boldsymbol{\varepsilon}})}{\mu_i}\left(e^{\theta r_{ji}}-1\right)\right\}.$$

Using the Markov inequality and the above, for any $j \in \mathcal{R}_k$ and $\theta \geq 0$, we obtain

$$\begin{aligned} &\mathbf{P}\left[\sum_{i\in\mathcal{O}_j} r_{ji}n_{i,\infty}^c > c\sum_{i\in\mathcal{O}_j}\frac{\lambda_i(u_{\text{ub},i}^*)r_{ji}}{\mu_i} - r_{jk}\right]\\ &\leq \mathbf{E}\left[e^{\sum_{i\in\mathcal{O}_j}\theta r_{ji}n_{i,\infty}^c}\right]\exp\left\{-\theta c\sum_{i\in\mathcal{O}_j}\frac{\lambda_i(u_{\text{ub},i}^*)r_{ji}}{\mu_i} + \theta r_{jk}\right\}\\ &= \exp\left\{c\sum_{i\in\mathcal{O}_j}\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta) + \theta r_{jk}\right\}\\ &= \exp\left\{\xi_{jk}^{\boldsymbol{\varepsilon}}(c,\theta)\right\}, \qquad (17) \end{aligned}$$

where $\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta)$ and $\xi_{jk}^{\boldsymbol{\varepsilon}}(c,\theta)$ were defined in (13) and (12), respectively. Optimizing the right hand side of (17) over all $\theta \geq 0$ to obtain the tightest bound yields

$$\begin{aligned} &\mathbf{P}\left[\sum_{i\in\mathcal{O}_j} r_{ji}n_{i,\infty}^c > c\sum_{i\in\mathcal{O}_j}\frac{\lambda_i(u_{\text{ub},i}^*)r_{ji}}{\mu_i} - r_{jk}\right]\\ &\qquad\leq \exp\left\{\inf_{\theta\geq 0}\xi_{jk}^{\boldsymbol{\varepsilon}}(c,\theta)\right\}. \qquad (18) \end{aligned}$$

Combining (18) with (15) and (16) yields

$$\mathbf{P}_{\text{loss}}^{k,c}(\mathbf{u}^{\boldsymbol{\varepsilon}}) \leq \sum_{j\in\mathcal{R}_k}\exp\left\{\inf_{\theta\geq 0}\xi_{jk}^{\boldsymbol{\varepsilon}}(c,\theta)\right\},$$

which establishes the bound in (11).

Let us now consider what happens as $c \to \infty$. For large c, $\xi_{jk}^{\boldsymbol{\varepsilon}}(c,\theta)$ will be dominated by $c\sum_{i\in\mathcal{O}_j}\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta)$. At $\theta = 0$,

$$\begin{aligned} \sum_{i\in\mathcal{O}_j}\Lambda_{ji}^{\boldsymbol{\varepsilon}}(0) &= 0,\\ \sum_{i\in\mathcal{O}_j}\left.\frac{\partial\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta)}{\partial\theta}\right|_{\theta=0} &= \sum_{i\in\mathcal{O}_j}\frac{r_{ji}(\lambda_i(u_i^{\boldsymbol{\varepsilon}}) - \lambda_i(u_{\text{ub},i}^*))}{\mu_i}. \end{aligned}$$

From Assumption A, for every $i \in \mathcal{O}$ and any $\boldsymbol{\varepsilon} > \mathbf{0}$ we have $\lambda_i(u_i^{\boldsymbol{\varepsilon}}) < \lambda_i(u_{\text{ub},i}^*)$. Furthermore, for every $i \in \mathcal{O}_j$, $r_{ji} > 0$. Therefore, $\sum_{i\in\mathcal{O}_j}\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta)$ achieves its minimum over $\theta \geq 0$ at some $\theta_j^*(\boldsymbol{\varepsilon}) > 0$ at which it holds $\sum_{i\in\mathcal{O}_j}\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta_j^*(\boldsymbol{\varepsilon})) < 0$. Note also that for all $j \in \mathcal{R}_k$

$$\inf_{\theta\geq 0}\xi_{jk}^{\boldsymbol{\varepsilon}}(c,\theta) \leq \left[c\sum_{i\in\mathcal{O}_j}\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta_j^*(\boldsymbol{\varepsilon})) + \theta_j^*(\boldsymbol{\varepsilon})r_{jk}\right],$$

and for large enough c the right hand side of the above is $O(c\sum_{i\in\mathcal{O}_j}\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta_j^*(\boldsymbol{\varepsilon})))$ which converges to $-\infty$ as $c \to \infty$. This establishes (14). ■

Remarks :

1. It should be noted that for small values of c the bound in (11) could be trivial, meaning that the right hand side might be larger than one.
2. As $c \to \infty$, however, the bound in (11) converges to zero exponentially fast like $\exp\{c\sum_{i\in\mathcal{O}_j}\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta_j^*(\boldsymbol{\varepsilon}))\}$, where $\theta_j^*(\boldsymbol{\varepsilon}) = \arg\inf_{\theta\geq 0}\sum_{i\in\mathcal{O}_j}\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta)$ and $\sum_{i\in\mathcal{O}_j}\Lambda_{ji}^{\boldsymbol{\varepsilon}}(\theta_j^*(\boldsymbol{\varepsilon})) < 0$.

We are now ready to state our asymptotic optimality result. We have seen that J_{ub}^c is linear in c. The optimal performance $J^{*,c}$ and the optimal performance J_{s}^c achieved by a static policy are also roughly linear in c. Thus, we will divide such quantities by c to make comparisons. The following theorem summarizes the result.

Theorem 3 *Consider either the revenue or the welfare maximization problem and assume that the functions $F_i(\lambda_i)$ are concave. Then,*

$$\lim_{c\to\infty}\frac{1}{c}J_{\text{s}}^c = \lim_{c\to\infty}\frac{1}{c}J^{*,c} = \lim_{c\to\infty}\frac{1}{c}J_{\text{ub}}^c.$$

Proof: To simplify the exposition we will provide the proof for the revenue maximization case; welfare maximization can be treated similarly. For some $\varepsilon > 0$, let $\boldsymbol{\varepsilon} = \varepsilon \mathbf{e}$, where $\mathbf{e}$ is the vector of all ones, and consider the static pricing policy $\mathbf{u}^{\boldsymbol{\varepsilon}} = \mathbf{u}^*_{\text{ub}} + \boldsymbol{\varepsilon}$. Let $J^c(\mathbf{u}^{\boldsymbol{\varepsilon}})$ be the resulting average revenue, which is no more than the optimal static revenue J^c_{s}. Thus,

$$\lim_{c\to\infty} \frac{1}{c} J^c_{\text{s}} \geq \lim_{c\to\infty} \frac{1}{c} J^c(\mathbf{u}^{\boldsymbol{\varepsilon}}) = \lim_{c\to\infty} \sum_{i=1}^{M} \lambda_i(u^{\boldsymbol{\varepsilon}}_i) u^{\boldsymbol{\varepsilon}}_i (1 - \mathbf{P}^{i,c}_{\text{loss}}(\mathbf{u}^{\boldsymbol{\varepsilon}})) = \sum_{i=1}^{M} \lambda_i(u^{\boldsymbol{\varepsilon}}_i) u^{\boldsymbol{\varepsilon}}_i.$$

In the last equality above we used Proposition 1 for all classes $i \in \mathcal{O}$, and the fact that for all $\varepsilon > 0$ demand is zero at $\mathbf{u}^{\boldsymbol{\varepsilon}}$ for all classes $i \notin \mathcal{O}$. Since the above inequality holds for any $\varepsilon > 0$, we take $\varepsilon \to 0$, which implies $\mathbf{u}^{\boldsymbol{\varepsilon}} \to \mathbf{u}^*_{\text{ub}}$ and, by the continuity of the demand functions,

$$\lim_{c\to\infty} \frac{1}{c} J^c_{\text{s}} \geq \sum_{i=1}^{M} \lambda_i(u^*_{\text{ub},i}) u^*_{\text{ub},i} = J^1_{\text{ub}}.$$

On the other hand, due to the suboptimality of the static policy and Theorem 2, $J^c_{\text{s}} \leq J^{*,c} \leq J^c_{\text{ub}} = cJ^1_{\text{ub}}$, and the result follows. ■

Theorem 3 establishes that in the limit $c \to \infty$ the upper bound of Theorem 2 is tight and the optimal solution of the upper bound problem, which is a static policy, is asymptotically optimal. Furthermore, Proposition 1 can be seen as characterizing the rate of convergence. This characterization allows us to determine how we should scale any given system to provide guarantees on the suboptimality gap of appropriately chosen static pricing policies. The following proposition describes the result. We state the result for the revenue maximization problem. It can be easily generalized to the welfare maximization problem as well.

Proposition 2 *Consider the revenue maximization problem and assume that the functions $F_i(\lambda_i)$ are concave. Let $\mathbf{u}^*_{\text{ub}}$ be the optimal solution to the upper bound problem of Theorem 2. For any $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_M) > \mathbf{0}$, consider the static policy $\mathbf{u}^{\boldsymbol{\varepsilon}} = \mathbf{u}^*_{\text{ub}} + \boldsymbol{\varepsilon}$, and let $J^c(\mathbf{u}^{\boldsymbol{\varepsilon}})$ its performance. For any given $\delta > 0$, let $(c^*, \boldsymbol{\varepsilon}^*)$ be an optimal solution of the following optimization problem*

$$\begin{aligned} \textit{min} \quad & c \\ \textit{s.t.} \quad & \sum_{i=1}^{M} \lambda_i(u^{\boldsymbol{\varepsilon}}_i) u^{\boldsymbol{\varepsilon}}_i \left(1 - \sum_{j\in\mathcal{R}_i} \exp\left\{ \inf_{\theta\geq 0} \xi^{\boldsymbol{\varepsilon}}_{ji}(c,\theta) \right\}\right) \\ & \qquad \geq \frac{\sum_{i=1}^{M} \lambda_i(u^*_{\text{ub},i}) u^*_{\text{ub},i}}{(1+\delta)} \\ & \boldsymbol{\varepsilon} \geq \mathbf{0}, \end{aligned} \tag{19}$$

where $\xi^{\boldsymbol{\varepsilon}}_{ji}(c,\theta)$ is defined in (12). Then, the performance of the static policy $\mathbf{u}^{\boldsymbol{\varepsilon}^*}$ in the c^*-scaled system satisfies

$$\frac{J^{*,c^*} - J^{c^*}(\mathbf{u}^{\boldsymbol{\varepsilon}^*})}{J^{c^*}(\mathbf{u}^{\boldsymbol{\varepsilon}^*})} \leq \delta. \tag{20}$$

Proof: Fix some $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_M) > \mathbf{0}$ and consider the static pricing policy $\mathbf{u}^{\boldsymbol{\varepsilon}}$ resulting in average revenue equal to $J^c(\mathbf{u}^{\boldsymbol{\varepsilon}})$. Due to the suboptimality of the static policy, Theorem 2, and Proposition 1, $J^c(\mathbf{u}^{\boldsymbol{\varepsilon}})$ satisfies

$$\sum_{i=1}^{M} \lambda_i(u^*_{\text{ub},i}) u^*_{\text{ub},i} = J^1_{\text{ub}} = \frac{1}{c} J^c_{\text{ub}} \geq \frac{1}{c} J^c(\mathbf{u}^{\boldsymbol{\varepsilon}}) = \sum_{i=1}^{M} \lambda_i(u^{\boldsymbol{\varepsilon}}_i) u^{\boldsymbol{\varepsilon}}_i (1 - \mathbf{P}^{i,c}_{\text{loss}}(\mathbf{u}^{\boldsymbol{\varepsilon}})) \geq \sum_{i=1}^{M} \lambda_i(u^{\boldsymbol{\varepsilon}}_i) u^{\boldsymbol{\varepsilon}}_i \left(1 - \sum_{j\in\mathcal{R}_i} \exp\left\{ \inf_{\theta\geq 0} \xi^{\boldsymbol{\varepsilon}}_{ji}(c,\theta) \right\}\right). \tag{21}$$

Using the same argument as in the proof of Theorem 3, we can first take $c \to \infty$ in the above and bring the blocking probabilities to zero. If we then take $\boldsymbol{\varepsilon} \to \mathbf{0}$ we conclude that the right hand side of (21) converges to its left hand side and the inequality is satisfied with equality. Thus, for any $\delta > 0$, we can find a scaling factor c and a static policy $\mathbf{u}^{\boldsymbol{\varepsilon}}$, such that $\frac{J^c_{\text{ub}} - J^c(\mathbf{u}^{\boldsymbol{\varepsilon}})}{J^c(\mathbf{u}^{\boldsymbol{\varepsilon}})} \leq \delta$, by solving the optimization problem in (19). More specifically, if $(c^*, \boldsymbol{\varepsilon}^*)$ is an optimal solution of (19) we have

$$\frac{J^{c^*}_{\text{ub}} - J^{c^*}(\mathbf{u}^{\boldsymbol{\varepsilon}^*})}{J^{c^*}(\mathbf{u}^{\boldsymbol{\varepsilon}^*})} \leq \delta,$$

and the desired result follows since $J^{c^*}_{\text{ub}} \geq J^{*,c^*}$. ■

A. *Structure of the Asymptotically Optimal Static Pricing Policy*

As we have seen the optimal solution of the upper bound problem of Theorem 2 provides a static pricing policy which is asymptotically optimal in the regime of many small users we considered. We will next characterize its structure. To that end, we will view the upper bound problem (9) as one involving optimization with respect to u_i, rather than λ_i. We will also write n_i in the form $\lambda_i(u_i)/\mu_i$. We start with the revenue maximization problem.

A.1 Revenue Maximization

The upper bound problem becomes

$$\begin{aligned} \max \quad & \sum_i \lambda_i(u_i) u_i \\ \text{s.t.} \quad & \sum_i \frac{\lambda_i(u_i) r_{ji}}{\mu_i} \leq C_j, \quad \forall j. \end{aligned} \tag{22}$$

Let $\mathbf{q} = (q_1, \ldots, q_L) \geq 0$ be the Lagrange multiplier vector, where q_j is associated with the capacity constraint on link j. Writing the problem in (22) as a minimization problem, its Lagrangean function becomes

$$\mathcal{L}(\mathbf{u},\mathbf{q}) = -\sum_{i=1}^{M}\lambda_i(u_i)u_i + \sum_{j=1}^{L} q_j\left(\sum_{i=1}^{M}\frac{\lambda_i(u_i)r_{ji}}{\mu_i} - C_j\right).$$

Assuming an interior solution $u_i \in (0, u_{i,\max})$, u_i should minimize

$$\left(-\lambda_i(u_i)u_i + \sum_j q_j\lambda_i(u_i)\frac{r_{ji}}{\mu_i}\right). \qquad (23)$$

Therefore, from the first order optimality condition we obtain

$$u_i = -\frac{\lambda_i(u_i)}{d\lambda_i(u_i)/du_i} + \sum_{j=1}^{L} q_j\frac{r_{ji}}{\mu_i}, \qquad \forall i. \qquad (24)$$

This structure is insightful. The first term is the reciprocal of the demand elasticity, prescribing that we should charge more to classes with more inelastic demand. The second term is a usage-based charge. Notice, that by complementary slackness conditions $q_j = 0$, if the corresponding constraint is not active, which can be interpreted as link j not being congested. On the other hand, if link j is congested (i.e., the corresponding constraint is satisfied with equality at the optimal solution), we charge each class a price $q_j > 0$ per unit of volume on link j. Here, we define as volume the quantity r_{ji}/μ_i, which is the bandwidth occupied times the expected holding time. Thus, the second term in (24) includes a charge for volume on congested links along the route $\mathcal{R}_i$ of class i.

This pricing structure is appealing from an implementation point of view. Large (backbone) networks might typically accommodate many service classes (number of offered services times number of origin-destination pairs), but consist of a relatively small number of links. Later on we will use this pricing structure and optimize over the shadow prices $\mathbf{q}$ to obtain near-optimal performance even away from the limiting regime.

A.2 Welfare Maximization

The case of welfare maximization can be treated similarly. Using (4) an interior solution $u_i \in (0, u_{i,\max})$ should minimize

$$\left(-\lambda_{0,i}\int_{u_i}^{u_{i,\max}} v f_i(v)dv + \sum_j q_j\lambda_i(u_i)\frac{r_{ji}}{\mu_i}\right),$$

which is analogous to the condition in (23) for the revenue maximization case. Therefore, from the first order optimality condition we obtain

$$\lambda_{0,i}\ u_i\ f_i(u_i) + \sum_j q_j\frac{r_{ji}}{\mu_i}\frac{d\lambda_i(u_i)}{du_i} = 0,$$

which, by using (3), becomes

$$u_i = \sum_{j=1}^{L} q_j\frac{r_{ji}}{\mu_i}, \qquad \forall i. \qquad (25)$$

As in revenue maximization, $q_j = 0$ for non-active constraints, thus, the pricing structure in (25) prescribes a usage-based charge for volume on all congested links along the route $\mathcal{R}_i$ of class i.

VII. Demand Substitution Effects

In this section we will extend the model we have considered so far to incorporate demand substitution effects. In particular, the model introduced in Section II assumes that the demand of each class $\lambda_i(u_i)$ is function of the price for that class only. We are interested in considering the situation where users might decide to use another class of service as a (non-perfect) substitute of their desired class if the latter one ends up being very expensive. Our main results so far extend to this situation as well. We will present a model that accounts for such substitution effects in Subsection VII-A. Following the development of the previous sections, we will develop an upper bound on the optimal performance in Subsection VII-B, establish the asymptotic optimality of static pricing in Subsection VII-C, and conclude this Section by characterizing the structure of the asymptotically optimal static policy in Subsection VII-D.

A. The Model

The model is in fact identical to the one introduced in Section II, with the exception that demand for each class i, $i = 1, \ldots, M$, is not only a function of u_i, but of the whole price vector $\mathbf{u}$, i.e., $\boldsymbol{\lambda}(\mathbf{u}) = (\lambda_1(\mathbf{u}), \ldots, \lambda_M(\mathbf{u}))$. We will maintain the rest of the notation that was introduced in Section II. We will denote the *offered load* on link j by

$$\rho_j(\mathbf{u}) \triangleq \sum_{i=1}^{M}\frac{r_{ji}\lambda_i(\mathbf{u})}{\mu_i C_j}, \qquad j = 1, \ldots, L.$$

We will be making the following assumption.

Assumption B
1. *If* $\lambda_i(\mathbf{u}) > 0$, *then* $\frac{\partial\lambda_i(\mathbf{u})}{\partial u_i} < 0$, *for* $i = 1, \ldots, M$;
2. $\frac{\partial\lambda_i(\mathbf{u})}{\partial u_k} \geq 0$, *for* $k \neq i$, $k = 1, \ldots, M$;
3. *if* $\lambda_i(\mathbf{u}) > 0$, *then* $\sum_{k=1}^{M}\frac{\partial\lambda_k(\mathbf{u})}{\partial u_i} < 0$, *for* $i = 1, \ldots, M$;
4. $\lim_{\mathbf{u}\to\infty}\lambda_i(\mathbf{u}) = 0$, *for* $i = 1, \ldots, M$, *where* $\mathbf{u} \to \infty$ *means* $\min_i u_i \to \infty$.

Assumption B-1 indicates that demand for any class is a strictly decreasing function of its own price. Assumption B-2 indicates that substitution among classes can take place, in the sense that the increase of the price for a class can increase the demand for other classes. Assumption B-3 states that only a fraction of demand lost for a class appears as demand for other classes (due to substitution). Assumption B-4 expresses the condition that as all prices increase, the demand will eventually decrease to zero for all classes.

As an example, the following linear demand functions satisfy Assumption B:

$$\begin{aligned}\lambda_1(\mathbf{u}) &= \lambda_{1,0} - \lambda_{1,1}u_1 + \lambda_{1,2}u_2,\\ \lambda_2(\mathbf{u}) &= \lambda_{2,0} + \lambda_{2,1}u_1 - \lambda_{2,2}u_2,\end{aligned} \tag{26}$$

for $\mathbf{u} \in \mathcal{U} = \{\mathbf{u} \mid \lambda_1(\mathbf{u}) \geq 0, \lambda_2(\mathbf{u}) \geq 0\}$, where $\lambda_{1,0}, \lambda_{2,0} > 0$, $\lambda_{1,1} > \lambda_{2,1} > 0$, $\lambda_{2,2} > \lambda_{1,2} > 0$.

Substitution effects can also be incorporated to our welfare maximization model of Section II-B. The model remains identical to the one introduced there with the exception that the user utility U_i of class i is a random variable depending on the whole price vector $\mathbf{u}$. In particular, we will assume that it has a probability density function, denoted by $f_i(u_i \mid u_j, j = 1, \ldots, M, j \neq i)$, conditional on the prices of all other classes. Potential calls decide to join the network if and only if the utility they extract exceeds the prevailing price. Thus, the actual arrival rate of class i calls under price $\mathbf{u}$ is

$$\lambda_i(\mathbf{u}) = \lambda_{0,i}\mathbf{P}[U_i \geq u_i \mid u_j, j = 1, \ldots, M, j \neq i],$$

where $\lambda_{0,i}$ is the peak class i demand (corresponding to zero prices in the revenue maximization model). A class i call joining the system extracts an expected utility equal to $\mathbf{E}[U_i \mid U_i \geq u_i;\ u_j, j = 1, \ldots, M, j \neq i]$, thus, social welfare for class i users is accumulated at a rate of

$$\lambda_i(\mathbf{u})\mathbf{E}[U_i \mid U_i \geq u_i;\ u_j, j = 1, \ldots, M, j \neq i].$$

Our objective remains to maximize the expected long-term average welfare rate, for which an expression can be written along the lines of (2).

Let us define the expected instantaneous rewards by $V_i(\mathbf{u}) \triangleq u_i$ and $V_i(\mathbf{u}) = \mathbf{E}[U_i \mid U_i \geq u_i;\ u_j, j = 1, \ldots, M, j \neq i]$ for the case of revenue and welfare maximization, respectively. We assume that $\boldsymbol{\lambda}(\mathbf{u})$ satisfies Assumption B in the welfare maximization case as well. Consequently, $\lambda_i(\mathbf{u})$ is non-decreasing in u_j for some $j \neq i$, which implies that $\mathbf{P}[U_i \geq u_i \mid u_j, j = 1, \ldots, M, j \neq i]$ is non-decreasing in u_j. We will be making the following assumption for the expected rewards.

Assumption C
For all $i = 1, \ldots, M$ and $\mathbf{u} \in \{\mathbf{u} \mid \lambda_i(\mathbf{u}) > 0\}$, $V_i(\mathbf{u})$ is a non-decreasing function of u_j for all $j \neq i$.

This assumption is trivially satisfied for the case of revenue maximization where $V_i(\mathbf{u}) = u_i$. For the case of welfare maximization it can be interpreted as follows. Each class i has a strong core constituency and can not be dominated by class j ($j \neq i$) customers who choose to use it as substitute when u_j increases. These "true" class i customers perceive that they are extracting a higher utility when the price of other services, u_j (for $j \neq i$), becomes relatively more expensive. Thus, $V_i(\mathbf{u})$ is non-decreasing in u_j for $j \neq i$. It also turns out $V_i(\mathbf{u})$ is non-decreasing in u_i. The next lemma establishes the result.

Lemma 1 *For all $i = 1, \ldots, M$ and $\mathbf{u} \in \{\mathbf{u} \mid \lambda_i(\mathbf{u}) > 0\}$, $V_i(\mathbf{u})$ is a non-decreasing function of u_i.*

Proof: The result is trivially true for the revenue maximization case where $V_i(\mathbf{u}) = u_i$. For welfare maximization we have

$$\begin{aligned}V_i(\mathbf{u}) &= \mathbf{E}[U_i \mid U_i \geq u_i;\ u_j, j = 1, \ldots, M, j \neq i]\\ &= \frac{\int_{u_i}^{\infty} v f_i(v \mid u_j, \forall j \neq i)\ dv}{\int_{u_i}^{\infty} f_i(v \mid u_j, \forall j \neq i)\ dv}.\end{aligned}$$

Taking the partial derivative we obtain

$$\frac{\partial V_i(\mathbf{u})}{\partial u_i} = \frac{f_i(u_i \mid u_j, \forall j \neq i) \int_{u_i}^{\infty} (v - u_i) f_i(v \mid u_j, \forall j \neq i)\ dv}{(\mathbf{P}[U_i \geq u_i | u_j, \forall j \neq i])^2},$$

which is clearly non-negative. ■

This lemma can be seen as expressing the fact that when u_i increases class i customers with relatively low utility for the service choose not to use it, thus, the ones that remain have higher utilities and drive $V_i(\mathbf{u})$ up.

B. An Upper Bound

We will follow the development of Section V. Assuming that the demand function is invertible, we can express the prices as a function of the arrival rates $\boldsymbol{\lambda}$; we will write $u_i(\boldsymbol{\lambda})$ for the class i price. Define $F_i(\boldsymbol{\lambda}) \triangleq \lambda_i u_i(\boldsymbol{\lambda})$ and $F_i(\boldsymbol{\lambda}) \triangleq \lambda_i \mathbf{E}[U_i \mid U_i \geq u_i(\boldsymbol{\lambda});\ u_j(\boldsymbol{\lambda}), j = 1, \ldots, M, j \neq i]$ for the case of revenue and welfare maximization, respectively. Assume that $F_i(\boldsymbol{\lambda})$ are concave functions of $\boldsymbol{\lambda}$ for all i. This is true, for example, for the demand functions in (26). The following result is analogous to Theorem 2. The proof is almost identical and is therefore omitted.

Theorem 4 *Consider the following nonlinear optimization problem*

$$\begin{aligned}\textit{maximize} \quad & \sum_{i=1}^{M} F_i(\boldsymbol{\lambda})\\ \textit{subject to} \quad & \lambda_i = n_i\mu_i, \qquad i = 1, \ldots, M,\\ & \sum_i n_i r_{ji} \leq C_j, \quad j = 1, \ldots, L,\end{aligned} \tag{27}$$

and let J_{ub} denote the optimal objective value. If $F_i(\boldsymbol{\lambda})$ is a concave function of $\boldsymbol{\lambda}$ for all $i = 1, \ldots, M$, then $J^* \leq J_{\mathrm{ub}}$.

C. Asymptotic Optimality of Static Pricing

We consider the same limiting regime of *"many small users"* of Section VI. We scale both demand and capacity by a scaling factor $c \geq 1$, while all other quantities are held fixed. Again, a superscript c will denote quantities of interest in the c-scaled system. The demand function becomes $\boldsymbol{\lambda}^c(\mathbf{u}) = (c\lambda_1(\mathbf{u}), \ldots, c\lambda_M(\mathbf{u}))$, the capacity of link j becomes cC_j, and the offered load on link j remains unchanged, i.e., $\rho_j^c(\mathbf{u}) = \rho_j(\mathbf{u})$, $j = 1, \ldots, L$. The normalized revenue or welfare maximization problem under a static pricing policy $\mathbf{u}$ can be formulated as:

$$\max_{\mathbf{u}\in\mathcal{U}^c} \frac{1}{c}\sum_{i=1}^{M} V_i(\mathbf{u})\lambda_i^c(\mathbf{u})\left(1-\mathbf{P}_{\text{loss}}^{i,c}(\mathbf{u})\right) = \max_{\mathbf{u}\in\mathcal{U}} \sum_{i=1}^{M} V_i(\mathbf{u})\lambda_i(\mathbf{u})\left(1-\mathbf{P}_{\text{loss}}^{i,c}(\mathbf{u})\right), \quad (28)$$

where $\mathcal{U}^c = \mathcal{U} = \{\mathbf{u} \mid u_i \geq 0, \lambda_i(\mathbf{u}) \geq 0, i = 1,\ldots,M\}$ is the feasible set for $\mathbf{u}$.

We use following asymptotic results for loss probabilities from [14]. As $c \to \infty$, under the static policy $\mathbf{u}$, the loss probability of each class converges to

$$\mathbf{P}_{\text{loss}}^{i,\infty}(\mathbf{u}) = 1 - \prod_{j=1}^{L}(1-B_j(\mathbf{u}))^{r_{ji}}, \quad i = 1,\ldots,M, \quad (29)$$

where $B_j(\mathbf{u}) \in [0,1)$, $j = 1,\ldots,L$, satisfy following conditions

$$\hat{\rho}_j(\mathbf{u}) \triangleq \sum_{i=1}^{M} \frac{r_{ji}\lambda_i^c(\mathbf{u})}{\mu_i\, cC_j}\prod_{l=1}^{L}(1-B_l(\mathbf{u}))^{r_{li}} = \sum_{i=1}^{M} \frac{r_{ji}\lambda_i(\mathbf{u})}{\mu_i C_j}\prod_{l=1}^{L}(1-B_l(\mathbf{u}))^{r_{li}} \begin{cases} = 1, & \text{if } B_j(\mathbf{u}) > 0, \\ \leq 1, & \text{if } B_j(\mathbf{u}) = 0. \end{cases} \quad (30)$$

Following Kelly [14], we will call $\hat{\rho}_j(\mathbf{u})$ the *reduced load* on link j. We can interpret these asymptotic results as follows. Calls are blocked independently at each link j in their route. In particular, class i demand is thinned by a factor of $(1-B_j(\mathbf{u}))^{r_{ji}}$ at link j and $\prod_{j=1}^{L}(1-B_j(\mathbf{u}))^{r_{ji}} = 1-\mathbf{P}_{\text{loss}}^{i,\infty}(\mathbf{u})$ can be seen as the proportion of accepted class i calls. This results into a satisfied demand for class i equal to $\lambda_i^c(\mathbf{u})\prod_{j=1}^{L}(1-B_j(\mathbf{u}))^{r_{ji}}$. We will use Kelly's [14] terminology and say that link j is *overloaded* if $B_j(\mathbf{u}) > 0$ (which implies $\hat{\rho}_j = 1$); if $B_j(\mathbf{u}) = 0$ we will say that it is *underloaded* ($\hat{\rho}_j < 1$) or *critically loaded* ($\hat{\rho}_j = 1$). We should note that although the conditions in (30) lead to unique values for the reduced loads $\hat{\rho}_j(\mathbf{u})$ and the loss probabilities $\mathbf{P}_{\text{loss}}^{i,\infty}(\mathbf{u})$, the parameters $B_j(\mathbf{u})$ might not have a unique value. In fact, the values of $B_j(\mathbf{u})$ are unique if the routing matrix $\mathbf{R}$ has rank L; otherwise, there exits a unique vector $(B_1(\mathbf{u}),\ldots,B_L(\mathbf{u}))$ with maximal support, i.e., a vector that solves (30) and maximizes the dimension of the set $\mathcal{B}(\mathbf{u}) \triangleq \{j \mid B_j(\mathbf{u}) > 0\}$. The following Lemma states an observation that would be useful later on.

Lemma 2 *The offered loads $\rho_j(\mathbf{u})$ satisfy $\rho_j(\mathbf{u}) \leq 1$ for all links $j = 1,\ldots,L$ if and only if $B_j(\mathbf{u}) = 0$ for all links $j = 1,\ldots,L$.*

Proof: We will first argue that if $\rho_j(\mathbf{u}) \leq 1$, for all j, then $B_j(\mathbf{u}) = 0$ for all j. Otherwise, suppose there is a link j with $B_j(\mathbf{u}) > 0$. Due to (30) there is at least one i for which $r_{ji}\lambda_i(\mathbf{u}) > 0$. Moreover, (30) also implies that $\hat{\rho}_j(\mathbf{u}) < \rho_j(\mathbf{u})$ and $\hat{\rho}_j(\mathbf{u}) = 1$. This contradicts the initial assumption $\rho_j(\mathbf{u}) \leq 1$. For the converse, note that if all links are either underloaded or critically loaded, i.e., $B_j(\mathbf{u}) = 0$ for all j, then $\rho_j(\mathbf{u}) = \hat{\rho}_j(\mathbf{u}) \leq 1$, for all j. ■

Another interesting observation is that due to (29), $B_j(\mathbf{u}) = 0$, for all j, implies that $\mathbf{P}_{\text{loss}}^{i,\infty}(\mathbf{u}) = 0$, for all $i = 1,\ldots,M$.

We next define the *normalized reward* of class i with respect to link j, for all links j with $r_{ji} > 0$, as follows

$$\hat{V}_{i,j}(\mathbf{u}) \triangleq \frac{V_i(\mathbf{u})}{r_{ji}/\mu_i}, \quad (31)$$

thus,

$$V_i(\mathbf{u})\lambda_i(\mathbf{u}) = \hat{V}_{i,j}(\mathbf{u}) \cdot \frac{r_{ji}\lambda_i(\mathbf{u})}{\mu_i} = \hat{V}_{i,j}(\mathbf{u})\hat{\lambda}_{i,j}(\mathbf{u}),$$

where $\hat{\lambda}_{i,j}(\mathbf{u}) \triangleq \frac{r_{ji}\lambda_i(\mathbf{u})}{\mu_i}$ is the *normalized demand* of class i for link j. We can interpret $\hat{V}_{i,j}(\mathbf{u})$ as reward per volume on link j, where volume has the same interpretation as in Section VI-A, that is, resource utilization times the expected holding time. For a given static pricing policy $\mathbf{u}$, the normalized rewards at link j are fixed and define an ordering among classes traversing link j. In particular, for any classes i and k traversing link j (i.e., $r_{ji}, r_{jk} > 0$), we will say that i *is more valuable* than k if $\hat{V}_{i,j}(\mathbf{u}) > \hat{V}_{k,j}(\mathbf{u})$. If calls occupy the *same* resource amount at all links in their route (i.e., for all i, $r_{ji} = r_i$ for all $j \in \mathcal{R}_i$ and $r_{ji} = 0$ for all $j \notin \mathcal{R}_i$), then the priority ordering of classes is the same on all links and $\hat{V}_{i,j}(\mathbf{u}) = \frac{V_i(\mathbf{u})}{r_i/\mu_i}$ define a unique priority ordering for the whole network. In this case, we will use the notation $\hat{V}_i(\mathbf{u}) \triangleq \frac{V_i(\mathbf{u})}{r_i/\mu_i}$. The following proposition is key in establishing our asymptotic optimality result.

Proposition 3 *Consider either the case of revenue or welfare maximization and assume that for any class $i = 1,\ldots,M$ and all links $j = 1,\ldots,L$*

$$r_{ji} = \begin{cases} r_i, & j \in \mathcal{R}_i, \\ 0, & \textit{otherwise.} \end{cases} \quad (32)$$

If $\mathbf{u}_{\text{s},\infty}$ solves the limiting case of problem (28), i.e.,

$$\max_{\mathbf{u}\in\mathcal{U}} \lim_{c\to\infty} \frac{1}{c}\sum_{i=1}^{M} V_i(\mathbf{u})\lambda_i^c(\mathbf{u})\left(1-\mathbf{P}_{\text{loss}}^{i,c}(\mathbf{u})\right) = \max_{\mathbf{u}\in\mathcal{U}} \sum_{i=1}^{M} V_i(\mathbf{u})\lambda_i(\mathbf{u})\prod_{j=1}^{L}(1-B_j(\mathbf{u}))^{r_{ji}}, \quad (33)$$

then

$$\rho_j(\mathbf{u}_{\text{s},\infty}) = \sum_{i \mid j\in\mathcal{R}_i} \frac{\lambda_i(\mathbf{u}_{\text{s},\infty})r_i}{\mu_i C_j} \leq 1, \quad j = 1,\ldots,L. \quad (34)$$

Proof: [1] The following discussion is about the lim-

[1]The present proof is a revised version of the proof that appeared

iting case $c \to \infty$. Let $\mathbf{u}$ be a static pricing policy such that on some links, the offered load is greater than 1. The average reward is

$$\sum_i V_i(\mathbf{u})\lambda_i(\mathbf{u}) \prod_l (1 - B_l(\mathbf{u}))^{r_{li}}, \tag{35}$$

where $B_l(\mathbf{u})$, $l = 1, \ldots, L$, satisfy (cf. (30))

$$\sum_i \frac{r_{ji}}{\mu_i} \lambda_i(\mathbf{u}) \prod_l (1 - B_l(\mathbf{u}))^{r_{li}} \le C_j, \quad j = 1, \ldots, L. \tag{36}$$

Consider the following linear programming (LP) problem with decision variables $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_M)$

$$\begin{aligned} \max \quad & \sum_i V_i(\mathbf{u})\lambda_i && (37) \\ \text{s.t.} \quad & \sum_i \frac{r_{ji}}{\mu_i}\lambda_i \le C_j, \quad j = 1, \ldots, L, \\ & 0 \le \lambda_i \le \lambda_i(\mathbf{u}), \quad i = 1, \ldots, M. \end{aligned}$$

Let $\tilde{\lambda}_i(\mathbf{u}) = \lambda_i(\mathbf{u}) \prod_l (1 - B_l(\mathbf{u}))^{r_{li}}$ for $i = 1, \ldots, M$. Clearly, $0 \le \tilde{\lambda}_i(\mathbf{u}) \le \lambda_i(\mathbf{u})$, which when combined with (36) implies that $\tilde{\boldsymbol{\lambda}}(\mathbf{u}) = (\tilde{\lambda}_1(\mathbf{u}), \ldots, \tilde{\lambda}_M(\mathbf{u}))$ is a feasible solution of (37). Due to (35), the objective value $\sum_i V_i(\mathbf{u})\tilde{\lambda}_i(\mathbf{u})$ is equal to the average reward of static policy $\mathbf{u}$.

Let $\hat{\boldsymbol{\lambda}}(\mathbf{u}) = (\hat{\lambda}_1(\mathbf{u}), \ldots, \hat{\lambda}_M(\mathbf{u}))$ denote an optimal solution of (37). Consider next the network under the static policy $\mathbf{u}$ but introduce a random admission control mechanism. More specifically, if $\lambda_i(\mathbf{u}) > 0$ then class i calls are accepted with a probability equal to $\hat{\lambda}_i(\mathbf{u})/\lambda_i(\mathbf{u})$. Otherwise, i.e., if $\lambda_i(\mathbf{u}) = 0$, $\hat{\lambda}_i(\mathbf{u}) = 0$ as well, and no admission control is applied. Thus, admitted class i calls arrive according to a Poisson process of rate $\hat{\lambda}_i(\mathbf{u})$, since their requests follow a Poisson process of rate $\lambda_i(\mathbf{u})$.

We will call System S_A the original one (without admission control), and System S_B the new system (with admission control). Note that in System S_B we have

$$\rho_j^B(\mathbf{u}) = \sum_i \frac{r_{ji}\hat{\lambda}_i(\mathbf{u})}{\mu_i C_j} \le 1, \qquad \forall j,$$

due to the feasibility conditions in (37). Thus, Lemma 2 implies that the blocking probabilities are equal to zero for all classes. The optimal value of problem (37) is simply the average reward in System S_B, thus, it is not less than the average reward in our original System S_A, which corresponds to the feasible solution $\tilde{\boldsymbol{\lambda}}(\mathbf{u})$.

Note that if $\hat{\lambda}_i(\mathbf{u}) = \lambda_i(\mathbf{u})$ for all i, then we can remove admission control. System S_B becomes identical to System S_A, which implies that the offered loads satisfy $\rho_j(\mathbf{u}) \le 1$ for all links j. Since we assumed that some links face an offered load higher than 1 under static policy $\mathbf{u}$, there exists a class i with $\hat{\lambda}_i(\mathbf{u}) < \lambda_i(\mathbf{u})$. Let us now construct a new policy $\mathbf{u}'$ as follows. Keep $u'_j = u_j$ for $j \ne i$, but increase the price u_i to a certain $u'_i > u_i$ such that under this new price vector $\mathbf{u}'$ we have $\lambda_i(\mathbf{u}') = \hat{\lambda}_i(\mathbf{u}) < \lambda_i(\mathbf{u})$. Such a policy $\mathbf{u}'$ exists due to Assumption B-1. Moreover, Assumption B-2 implies that $\lambda_j(\mathbf{u}') \ge \lambda_j(\mathbf{u})$ for $j \ne i$.

Replace now $\mathbf{u}$ with $\mathbf{u}'$ in the optimization problem (37). It can be seen that $\hat{\boldsymbol{\lambda}}(\mathbf{u})$ is a feasible solution to the new problem. Due to Assumption C and Lemma 1, $\sum_{i=1}^M V_i(\mathbf{u}')\hat{\lambda}_i(\mathbf{u})$ is no less than $\sum_{i=1}^M V_i(\mathbf{u})\hat{\lambda}_i(\mathbf{u})$, which is the average reward of System S_B.

Let $\boldsymbol{\lambda}^*(\mathbf{u}')$ be an optimal solution of the new optimization problem (37) corresponding to the new price vector $\mathbf{u}'$. A new System S_C under policy $\mathbf{u}'$ and with admission control can be constructed similarly to System S_B. The average reward in System S_C is no less than System S_B, therefore, no less than the original System S_A.

Again we check if we can remove the admission control in System S_C. If not, we repeat the above procedure until we find a static policy such that the offered loads are no greater than 1 on all links. Throughout, the average reward does not decrease. Assumptions B-3 and B-4 guarantee the existence of such a static policy.

To summarize, we started from an arbitrary price vector $\mathbf{u}$ under which some links have offered loads greater than one and constructed a price vector with higher average reward and offered loads no greater than 1 on all links j. We conclude that the optimal limiting static policy $\mathbf{u}_{s,\infty}$ must satisfy $\rho_j(\mathbf{u}_{s,\infty}) \le 1$ for all links j. ∎

Due to Lemma 2, the result of Proposition 3 implies that at the optimal static prices in the limiting regime, $\mathbf{u}_{s,\infty}$, all links in the network are underloaded or critically loaded, and all classes experience zero blocking probabilities. The following theorem is an immediate consequence of these observations and Proposition 3.

Theorem 5 *Consider either the case of revenue or welfare maximization and assume that for any class $i = 1, \ldots, M$ and all links $j = 1, \ldots, L$*

$$r_{ji} = \begin{cases} r_i, & j \in \mathcal{R}_i, \\ 0, & \text{otherwise.} \end{cases}$$

The optimal static policy in the limiting regime, $\mathbf{u}_{s,\infty}$, solves the following optimization problem:

$$\begin{aligned} \textit{maximize} \quad & \sum_{i=1}^M V_i(\mathbf{u})\lambda_i(\mathbf{u}) && (38) \\ \textit{subject to} \quad & \sum_{i=1}^M \frac{\lambda_i(\mathbf{u})r_{ji}}{\mu_i} \le C_j, \qquad j = 1, \ldots, L. \end{aligned}$$

The optimization problem in (38) is in fact the same as the upper bound problem in (27) (with the exception that decision variables are the prices instead of the arrival rates). Thus, in the limiting regime ($c \to \infty$), the optimal static

in print which contained an error. In this new proof we do not use the notion of normalized reward defined earlier. Moreover, the assumption on resource requirements in Eq. (32) is no longer necessary. We maintain it in order to keep the statement of the Proposition identical with the one that appeared in print.

policy achieves the upper bound and it is *asymptotically optimal*.

D. Structure of the Asymptotically Optimal Static Pricing Policy

As in Section VI-A, where we considered the original model of Section II, we will next characterize the structure of asymptotically optimal prices for the modified model of Section VII-A.

Let us first focus on the revenue maximization case. Consider the problem in (38) and rewrite it as a minimization problem. Let $\mathbf{q} = (q_1, \ldots, q_L) \geq 0$ be the Lagrange multiplier vector, where q_j is associated with the capacity constraint on link j. The Lagrangean function becomes

$$\mathcal{L}(\mathbf{u}, \mathbf{q}) = -\sum_{i=1}^{M} \lambda_i(\mathbf{u}) u_i + \sum_{j=1}^{L} q_j \left(\sum_{i=1}^{M} \frac{\lambda_i(\mathbf{u}) r_{ji}}{\mu_i} - C_j \right). \tag{39}$$

Assuming an interior solution, $\mathbf{u}$ should satisfy

$$\nabla \boldsymbol{\lambda}(\mathbf{u})\mathbf{u} = -\boldsymbol{\lambda}(\mathbf{u}) + \sum_{j=1}^{L} q_j \sum_{i=1}^{M} \frac{r_{ji}}{\mu_i} \nabla \lambda_i(\mathbf{u}), \tag{40}$$

where $\nabla \boldsymbol{\lambda}(\mathbf{u})$ is the gradient of the vector function $\boldsymbol{\lambda}(\mathbf{u})$, i.e., an $M \times M$ matrix with (i,j) element equal to $\frac{\partial \lambda_j(\mathbf{u})}{\partial u_i}$.

Welfare maximization can be treated similarly. One can write down the optimality conditions for the problem in (38) and solve them analytically for relatively simple forms of the utility density functions $f_i(\cdot)$. The structure of those conditions is rather complex, so one would have to resort to numerical solution methods for the general case.

VIII. Large Scale Problems

In this section we discuss how the pricing policies we have considered so far can be computed and applied to large scale systems.

Large networks consist of numerous classes (equal to the number of offered services times the origin-destination pairs) and many links with large capacities. As a result, the state space $\mathcal{S} = \{\mathbf{n} \mid \mathbf{Rn} \leq \mathbf{C}\}$ becomes enormous and it is intractable to compute the optimal (dynamic) policy. One could potentially leverage recent approximate dynamic programming techniques to compute an approximately optimal dynamic policy. This direction has been successfully explored in [13] and can be generalized in the network setting. The sheer dimensionality of the network problem though, makes the computational effort more challenging.

In this paper we are focusing on static pricing policies because they are simpler and have significant implementation advantages over dynamic ones; we have outline those in the Introduction. As we commented in Section IV, computing the optimal static policy exactly is also computationally intractable. Instead we will experiment with the following two approaches to compute effective static pricing policies:

1. **Policy from the Upper Bound**. As we have seen the optimal solution of the upper bound problem in (9) forms a static pricing policy for our original model of Section II. For the system with demand substitution effects of Section VII-A a static policy is formed by the optimal solution of the upper bound problem in (27). In both cases, we have seen that in the limiting regime of many small users these policies are asymptotically optimal. Furthermore, they are quite easy to obtain; their computation amounts to solving a nonlinear programming problem with $O(L)$ linear constraints and $O(M)$ decision variables; for which effective algorithms exist.

2. **Using the structure of asymptotically optimal static policy**. A concern with the static policy from the upper bound is that it might not perform as well away from the limiting regime. Some earlier experience with the single node problem in [13] indicates that the structure of the asymptotically optimal static policy is effective away from the limiting regime but the values of the various parameters might not be appropriate away from the limit. More specifically, note that the structure of the policies in Sections VI-A and VII-D depends on the selection of a set of shadow prices (Lagrange multipliers) for the resources at all congested links of the network. To improve upon the policy obtained from the upper bound we seek to optimize the performance objective (revenue or social welfare) over those shadow prices. To that end, we will employ a simulation-based method.

In particular, we adopt the structure of the policies of Sections VI-A and VII-D and view them as functions of the Lagrange multipliers q_j, $j = 1, \ldots, L$. During the course of a simulation of the system we obtain "gradient information" which is used to optimize over q_j's. To this end, we will apply a technique developed by Marbach and Tsitsiklis [19]. Under some technical assumptions, the convergence of the algorithm to a stationary point (i.e., a point where the gradient of the performance with respect to $\mathbf{q}$ is zero) can be guaranteed (w.p. 1). Alternatively, we could optimize over prices directly, but the dual approach of optimizing over the Lagrange multipliers q_j is more preferable in large networks since, typically, the number of classes is much larger than the number of links.

IX. Numerical Results

In this section we tackle, numerically, some illustrative network pricing problems using the ideas discussed in the previous sections. We will present revenue maximization problems. The qualitative conclusions would not be much different in welfare maximization. We start with a network that conforms to the model of Section II.

A. Networks without Substitution Effects

Our first example, depicted in Figure 1, is a large (backbone) network. It consists of 9 nodes, 13 links, and provides 59 classes of services, the parameters of which are omitted in the interest of space. [2]

Table I compares the upper bound J_{ub} (cf. Thm. 2) with the two policies proposed in Section VIII. We use $J(\mathbf{u}^*_{\text{ub}})$ and J_{sim} to denote the performance of the policy obtained

[2]Service parameters and link capacities are available online at `http://ionia.bu.edu/`

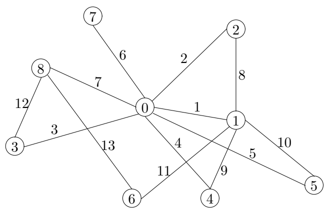

Fig. 1. A 9-node, 13-link network with 59 service classes.

from the upper bound problem and the simulation-based optimization approach, respectively. We observe that J_{sim} is quite close to the optimal. Note that the percentage gap in Table I is an upper bound on the suboptimality gap.

J_{ub}	$J(\mathbf{u}^*_{\text{ub}})$	J_{sim}	$\frac{J_{\text{ub}}-J_{\text{sim}}}{J_{\text{ub}}} \times 100\%$
12597.6	12117.4	12209.6	3.08%

Table I
Comparing the various policies for the network of Figure 1.

It is perhaps of interest to use Proposition 2 to compute by how much we should scale the network to achieve a given suboptimality gap. Using the notation introduced there, a suboptimality gap of $\delta = 0.1$ is guaranteed by scaling the network by $c = 3.30$ and using policy $\mathbf{u}^{\boldsymbol{\varepsilon}}$ with $\boldsymbol{\varepsilon} = 0.59\mathbf{e}$, where $\mathbf{e}$ is the vector of all ones. Similarly, $\delta = 0.05$ is achieved with $c = 10.74$ and $\boldsymbol{\varepsilon} = 0.38\mathbf{e}$. Finally, $\delta = 0.01$ is achieved with $c = 218.07$ and $\boldsymbol{\varepsilon} = 0.12\mathbf{e}$. Note that for simplicity of the calculations involved we only considered $\boldsymbol{\varepsilon} = \epsilon\mathbf{e}$ in the optimization problem (19). The results can be improved by considering arbitrary $\boldsymbol{\varepsilon}$. Clearly, these guarantees come from (crude) bounds on the blocking probability and are not meant to be very tight. Our optimized policy ($\mathbf{u}^*_{\text{sim}}$), for example, would be much closer to optimal in each of those scaled systems. Nevertheless, Proposition 2 provides a simple way to quickly assess efficiency gains by scaling the system.

B. Networks with Substitution Effects

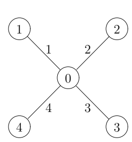

Fig. 2. A 5-node, 4-link network with 12 service classes.

We next consider the network in Figure 2, and incorporate demand substitution effects. The network provides 12 classes of service [3]; classes 1 and 7, and 2 and 8, can be used as substitutes of each other. Table II compares the upper bound J_{ub} (cf. Thm. 4) with the two policies proposed in Section VIII.

J_{ub}	$J(\mathbf{u}^*_{\text{ub}})$	J_{sim}	$\frac{J_{\text{ub}}-J_{\text{sim}}}{J_{\text{ub}}} \times 100\%$
814.84	765.61	783.85	3.8%

Table II
Comparing the various policies for the network of Figure 2.

As in the previous example, we conclude that the optimized version (via the simulation-based optimization approach) of our asymptotically optimal static pricing policy is reasonably close to the optimal.

[3]Again, see `http://ionia.bu.edu/` for service parameters and link capacities.

X. Conclusions

We considered a loss network model and studied the problem of pricing the use of the available resources under both revenue and welfare maximization objectives. Our results generalize [13] in several directions. We established that static pricing is asymptotically optimal in a regime of many small users. To that end, we showed that in this limit the blocking probabilities under an appropriate static pricing policy converge to zero exponentially fast. We characterized this exponential rate of convergence, which allowed us to obtain simple estimates on the size of the network in which static pricing is within a given distance from the optimal.

We also considered an extension of our original demand model that incorporates demand substitution effects among various classes. Using a different set of techniques, we established that our main asymptotic optimality result of static pricing holds for that model as well.

For both demand models and under both objectives, we characterized the structure of asymptotically optimal static prices and used this structure to obtain near-optimal policies away from the limiting regime. To that end, we employed a simulation-based optimization method that optimizes policy parameters by obtaining gradient information throughout the course of a simulation of the system. Our approach can handle large, realistic, problem sizes. Admittedly, the simulation-based algorithm needs global information on the state of the network to converge to the optimal, which is not appealing in practice. It is of interest to develop asynchronous and distributed optimization algorithms, as for example is done in Low and Lapsley [20] for a different (flow control) problem.

In practice, where demand is nonstationary but slowly varying, the policies we proposed lead to *time-of-day pricing*. There is substantial accumulated experience with such policies in the telecommunications industry, which facilitates their actual implementation. A practical implementation would also need to be coupled with a demand estimation mechanism (in fact, only demand elasticity information is needed). The proposed simulation-based optimization approach can be driven by the actual operation of the network, instead of a simulation. In this setting, a demand estimation mechanism can be naturally be incorporated.

Acknowledgments

We would like to thank John Tsitsiklis and David Gamarnik for useful discussions on these topics. We also thank for comments seminar participants at the NSF PI

Workshop in Networking (UC Irvine, November 2000) and at the IBM T.J. Watson Research Center (March 2001).

References

[1] J. K. MacKie-Mason and H. R. Varian, "Pricing the Internet," in *Public Access to the Internet*, B. Kahin and J. Keller, Eds. Prentice-Hall, 1994.

[2] F. P. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.

[3] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[4] R.J. Gibbens and F.P. Kelly, "Resource pricing and the evolution of congestion control," *Automatica*, vol. 35, no. 12, pp. 1969–1985, 1999.

[5] R.J. La and V. Anantharam, "Charge-sensitive TCP and rate control in the Internet," Preprint, 2000, available at `http://divine.eecs.berkeley.edu/~ananth/`.

[6] S. Kunniyur and R. Srikant, "End-to-end congestion control: utility functions, random losses and ECN marks," in *Proceedings of the INFOCOM conference*, Tel Aviv, Israel, 2000.

[7] S. H. Low, "Equilibrium bandwidth and buffer allocations for elastic traffics," *IEEE/ACM Transactions on Networking*, vol. 8, no. 3, pp. 373–383, 2000.

[8] F. P. Kelly, "Notes on effective bandwidths," in *Stochastic Networks: Theory and Applications*, S. Zachary, I.B. Ziedins, and F.P. Kelly, Eds., vol. 9, pp. 141–168. Oxford University Press, 1996.

[9] I. Ch. Paschalidis, "Class-specific quality of service guarantees in multimedia communication networks," *Automatica, (Special Issue on Control Methods for Communication Networks), Anantharam and Walrand Eds.,*, vol. 35, no. 12, pp. 1951–1968, 1999.

[10] F. P. Kelly, "On tariffs, policing and admission control for multiservice networks," *Operations Research Letters*, vol. 15, pp. 1–9, 1994.

[11] C. Courcoubetis, F.P. Kelly, and R. Weber, "Measurement-based usage charges in communications networks," *Operations Research*, vol. 48, no. 4, pp. 535–548, 2000.

[12] G. Gallego and G. van Ryzin, "A multiproduct dynamic pricing problem and its applications to network yield management," *Operations Research*, vol. 45, no. 1, pp. 24–41, 1997.

[13] I. Ch. Paschalidis and J. N. Tsitsiklis, "Congestion–dependent pricing of network services," *IEEE/ACM Transactions on Networking*, vol. 8, no. 2, pp. 171–184, 2000.

[14] F. P. Kelly, "Loss networks," *The Annals of Applied Probability*, vol. 1, pp. 319–378, 1991.

[15] K.W. Ross, *Multiservice loss models for broadband telecommunication networks*, Springer-Verlag, 1995.

[16] C. A. Courcoubetis and M. I. Reiman, "Pricing in a large single link loss system," in *Teletraffic Engineering in a Competitive World*, P. Key and D. Smith, Eds., pp. 737–746. Elsevier, 1999.

[17] D.P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. I and II, Athena Scientific, Belmont, MA, 1995.

[18] G.M. Louth, *Stochastic Networks: Complexity, Dependence and Routing*, Ph.D. thesis, University of Cambridge, Cambridge, U.K., 1990.

[19] P. Marbach and J.N. Tsitsiklis, "Simulation-based optimization of Markov reward processes," *IEEE Transactions on Automatic Control*, vol. 46, no. 2, pp. 191–209, 2001.

[20] S.H. Low and D.E. Lapsley, "Optimization flow control, I: Basic algorithm and convergence," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 861–874, 1999.

Ioannis Ch. Paschalidis(M'96) was born in Athens, Greece, in 1968. He received the Diploma degree in Electrical and Computer Engineering from the National Technical University of Athens, Athens, Greece, in 1991, and the S.M. and Ph.D. degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, in 1993 and 1996, respectively. During the summer of 1996 he was a Post-Doctoral Associate at the Laboratory for Information and Decision Systems, MIT, and since September 1996 he has been with Boston University where he currently is Associate Professor of Manufacturing Engineering. His current research interests include the analysis and control of stochastic systems, large deviations theory, queueing theory, optimization, pricing, and revenue management. The main application areas he is targeting include communication networks and systems, manufacturing systems, supply chains, and distribution systems.

He has received an NSF CAREER award (2000), and the second prize in the 1997 George E. Nicholson paper competition by INFORMS. He has served in the program committees of several conferences, including, the INFORMS Applied Probability Conference, the IEEE Conference on Decision and Control, and the INFOCOM. He is an associate editor of Operations Research Letters.

Yong Liu (SM'00) was born in Xianyang, China, in 1973. He received his B.E. in Information and Control Engineering and an M.E. in Systems Engineering both from the Xi'an Jiaotong University, China, in 1994 and 1997, respectively. He is currently a Ph.D. candidate at Boston University. His research interests include analysis and optimization of manufacturing systems and communication systems, pricing and revenue management.