# Combined Economic Modeling and Traffic Engineering: Joint Optimization of Pricing and Routing in Multi-Service Networks

Debasis Mitra[a] K. G. Ramakrishnan[a] Qiong Wang[a]

[a]Bell Labs, Lucent Technologies, Murray Hill, NJ 07974,
mitra,qwang@research.bell-labs.com

We consider a network, with fixed topology and link bandwidths, that offers multiple services, such as voice and data, each having characteristic price elasticity to demand, and quality of service and policy requirements on routing. We develop a revenue maximization model in which pricing and routing are jointly optimized. In this model, prices, which depend on service type and origin-destination, determine demands, that are routed, subject to their constraints, so as to maximize revenue, which is earned when the demand is delivered to its destination. The model is flow-based. We study the basic properties of the optimal solution and prove that link shadow costs provide the basis for selecting optimal prices and determining optimal routing policies. We present illustrative numerical examples in which network bandwidth is scaled to grow.

## 1. INTRODUCTION

Pricing and routing are two basic mechanisms for managing network resources. In our model, price determines demand, as well as the revenue that is generated by satisfying each unit of demand. Both roles are important in revenue maximization, which is the objective in this work. For revenue to be earned, demands generated have to be matched to the fixed network resources. Consequently, the optimal demand generation for the various interacting multi-service traffic streams, as well as the allocation of network resources to these streams, is strongly dependent on their unit prices. Routing is the mechanism for allocating network resources, i.e., bandwidth in network links, to the traffic streams. These relationships call for a combined study in which the interaction between pricing and routing is taken into account. With this as motivation, we develop, solve, and analyze a joint pricing and routing optimization model in this paper.

Traditionally, network pricing and routing have been studied as separate problems. Pricing schemes have typically been developed under the abstraction of a single-link network, in which routing is trivialized. Most routing and traffic engineering studies, on the other hand, take demand as exogenously given, which effectively makes pricing irrelevant. See, for instance, [3], [13], [9], and references therein. See also [12], [8], and [17], for congestion pricing in networks, which is fundamentally separated from the subject of this paper by issues of demand modeling and time scales; however, there are certain common underpinnings, such as shadow costs.

The presence of multiple services is a major influence in our analysis. Different services have different price-demand functions, characterized by distinct demand elasticity to price. Also, each service has its own Quality of Service (QoS) and policy requirements on the routing. Consequently, their prices as well as their respective representations in the total traffic are typically different. Prices and routes for each service are selected to maximize total network revenue.

The connection between pricing and routing is established through link shadow costs, which are the Lagrange multipliers corresponding to the link capacity constraints. We prove the important result that in the optimal solution, these quantities should be used both as the basis for setting optimal prices and for routing demands between node-pairs. We use this connection to demonstrate that the optimal pricing strategy restricts the optimal routes to a small set of least-cost routes. We further give a procedure that, without causing any reduction in the generated revenue, gives simplified routing implementations that minimize traffic splitting between alternate routes. Finally, we develop numerical case studies for inferring properties of optimal solutions for large networks, in which link bandwidths are scaled to grow.

The organization of the paper is as follows: Section 2 gives some essential background for the economic modeling and traffic engineering. Section 3 develops the joint optimization model and discusses basic characteristics of the optimal solution. Section 4 gives the results from the analyses of optimal routing for optimal pricing. Section 5 discusses properties of the optimal solution through case studies. Section 6 concludes the paper.

## 2. PRELIMINARIES IN ECONOMIC MODELING AND TRAFFIC ENGINEERING

Since these concepts are used extensively in subsequent sections, we briefly describe here, first, models of constant price elasticity of demands for voice, data, and other services, and second, admissible route sets in traffic engineering. This material is from [10] and [13], respectively.

### 2.1. Constant Price-elasticity Demand Models for Bandwidth

Forecasting demand is notoriously difficult, especially during periods of unprecedented transformation. A well-known illustration is the growth of the Internet, which has consistently been underestimated. Yet the relationship between price and demand, as given by constant price elasticity demand models, has a credible record that extends to several industries and time periods that cover several decades. A well-known example is Moore's Law, which is widely accepted in the microprocessor and dynamic random access memories (DRAM) industries. For microprocessors, the law states that the speed of chips double every 18 months while price remains constant. The availability of higher computer speeds and lower prices per unit speed generate new applications and markets that were previously impractical. The net effect is that the demand for chips more than doubles every 18 months, i.e., remains ahead of the growth in processor speed.

To capture these properties in a demand model, we use the concept of price elasticity. Let $D$ denote demand and $P$ denote price for a particular service. Consider demand as a function of price, i.e. $D = D(P)$. The price elasticity of demand, $\epsilon$, associated with a fixed given time-interval is defined as the negative ratio of the relative change in demand

and the relative change in price during that interval, i.e.,

$$\epsilon = -\frac{\Delta D/D}{\Delta P/P}, \tag{1}$$

where $\Delta D$ and $\Delta P$ denote the changes in demand and price. In the typical situation when demand increases as price decreases, the value of $\epsilon$ defined in (1) is positive. If $\epsilon > 1$, then any relative reduction in price leads to a larger relative increase in demand.

Taking the limit of (1) as the interval of change in price becomes infinitesimal, we obtain:

$$\epsilon = -(\frac{P}{D})\frac{dD}{dP}. \tag{2}$$

Since revenue, denoted by $R$, is given by the product of price and demand, i.e. $R = PD$, we obtain from (2) that:

$$\frac{dR/dP}{R} = -(\frac{\epsilon - 1}{P}). \tag{3}$$

Hence depending on whether $\epsilon > 1$, $\epsilon = 1$, or $\epsilon < 1$, reduction in price leads to increase, no change, or reduction in revenue.

A demand function with constant price elasticity is obtained by assuming that (2) holds for all possible $P$ and $D$ with $\epsilon$ held fixed. Hence, for $P > 0$,

$$D = \frac{A}{P^{\epsilon}}. \tag{4}$$

The constant $A$ is equal to the value of $D$ when $P = 1$, and it may be interpreted as a demand potential.

Earlier studies of the DRAM industry indicate that the price-demand relationship is well characterized by a constant price-elasticity function[10]. Techniques for estimating elasticity may be found in [6]. The traditional elasticity estimate for voice traffic is approximately 1.05, though France Telecom obtained a value as high as 1.337[2]. Lanning et. al[10] provide elasticity data for a variety of telecom equipment, and also a methodology for inferring elasticity of data services. See also [11]. The estimate of elasticity for data services given in [10] is in the range of 1.3-1.7.

We conclude by noting that the analytic framework of this paper is sufficiently general to accommodate demand models other than constant price elasticity models. However, constant price elasticity models appear to be sensible and appropriate for bandwidth.

### 2.2. Admissible Route Sets

There are two concepts that we rely upon for the planning of the delivery of QoS, namely, effective bandwidth of individual flows, or connections, which depends on the connection's service type, and admissible route sets. For service $s$ and (source, destination) pair $(\sigma_1, \sigma_2)$, the total bandwidth for $(s, (\sigma_1,\sigma_2))$ is given by the product of the effective bandwidth of individual connections of the service and the rate of connections. (We abbreviate $(\sigma_1,\sigma_2)$ to $\sigma$, and refer to $(s,\sigma)$ as a *stream*). This paper deals with the total stream bandwidth, and consequently the notion of effective bandwidth, while obviously important, does not feature directly in the analysis.

Admissible route sets, on the other hand, are ubiquitous in this paper. It is a device that allows us to take into account constraints on routing imposed by QoS and, more generally, policy considerations. For instance, converged data networks will support high quality voice calls, as well as packet video and other delay sensitive services, in addition to data in a variety of service offerings. To handle their end-to-end constraints, we let $\mathcal{R}(s,\sigma)$ denote the set of admissible route for the stream $(s,\sigma)$. For example, real-time services, such as voice and video, may require routes with lengths not exceeding specified thresholds, on account of propagation delay, and typically, restrictions will also apply on the number of hops in the routes, since each hop is associated with an additional switching node and consequent incremental delay. In contrast, for delay insensitive service classes, the distance and hop constraints would be less stringent. Importantly, the admissibility of a route will also depend on policy, which might reflect diverse considerations, such as security, the capability of switching nodes in the routes to handle certain services, link capacity, etc. Generating the admissible route sets, $\mathcal{R}(s,\sigma)$, is a substantial task in itself. In this paper, we consider these sets are given.

In Section 5 where we present results from our numerical investigations, the only restriction that the admissible routes are required to satisfy is on the maximum number of hops, a service-specific parameter. This criterion is just for illustrative purposes.

We make the simplifying assumption that for $\sigma_1$ and $\sigma_2$ which are adjacent, i.e., a direct link connects $\sigma_1$ to $\sigma_2$ , the admissible route set $\mathcal{R}(s,\sigma)$ always includes the direct link for at least one service. That is, the direct link is always an admissible route for its end points for at least one service.

## 3. PROBLEM FORMULATION AND BASIC SOLUTION PROPERTIES

In this section, we formulate a revenue maximization problem for a service provider's network model, which incorporates constant price elasticity demand models for each of several services. The joint optimization is with respect to price and routing decisions for each stream in the network. Next, we follow Lagrange's method for concave programming, and show that the shadow costs are effective devices for traffic engineering. These shadow costs are the bases for the optimum "minimum cost" routing policy.

### 3.1. Model and Revenue Maximization

First we review the notation. Let $s$ $(s = 1, 2 \ldots, S)$ denote service class. Let $(\sigma_1, \sigma_2)$ denote (source, destination) pair, and $(s, \sigma)$ a stream. $P_{s\sigma}$ is the price for unit bandwidth for stream $(s, \sigma)$. $D_{s\sigma}$ is the bandwidth demand for stream $(s, \sigma)$. The (fixed) capacity, in units of bandwidth, of link $l$, is denoted by $C_l$ $(l = 1, 2, \ldots, L)$. For $r \in \mathcal{R}(s, \sigma)$, we call $(s, r)$ a service route, and let $X_{sr}$ denote the carried bandwidth on this service route. For each $(s, \sigma)$,

$$D_{s\sigma} = \frac{A_{s\sigma}}{P_{s\sigma}^{\epsilon_{s\sigma}}} \quad \text{or} \quad P_{s\sigma} = (\frac{A_{s\sigma}}{D_{s\sigma}})^{1/\epsilon_{s\sigma}} \tag{5}$$

where $\epsilon_{s\sigma}$ and $A_{s\sigma}$ are respectively the elasticity and demand potential for the stream. As previously noted, we assume that $\epsilon_{s\sigma} > 1$, for all $(s, \sigma)$.

The revenue generated on a stream depends on the product of the carried bandwidth on the stream and the unit price for bandwidth for the stream. The network revenue, $W$, is obtained by summing over all streams.

$$W = \sum_{s,\sigma} P_{s\sigma} \sum_{r \in \mathcal{R}(s,\sigma)} X_{sr} \tag{6}$$

The revenue maximization problem is:

$$\max_{\{P_{s\sigma}\},\{X_{sr}\}} W \tag{7}$$

subject to the following constraints:

$$\sum_{r \in \mathcal{R}(s,\sigma)} X_{sr} \leq D_{s\sigma} \qquad \forall (s,\sigma) \tag{8}$$

$$\sum_{(s,\sigma)} \sum_{r \in \mathcal{R}(s,\sigma): l \in r} X_{sr} \leq C_l \qquad \forall l \tag{9}$$

$$P_{s\sigma} \geq 0 \qquad\qquad \forall (s,\sigma) \tag{10}$$

$$X_{sr} \geq 0 \qquad \forall (s,\sigma), \forall r \in \mathcal{R}(s,\sigma) \tag{11}$$

In the above formulation, the price and demand variables are both represented, even though these variables are related through the demand model in (5). We will find it convenient to eliminate the price variables and retain only the demand variables. The optimum prices may, of course, be recovered from the demands in the solution. Transforming the objective function gives:

$$\max_{\{D_{s\sigma}\},\{X_{sr}\}} W = \sum_{(s,\sigma)} (\frac{A_{s\sigma}}{D_{s\sigma}})^{1/\epsilon_{s\sigma}} \sum_{r \in \mathcal{R}(s,\sigma)} X_{sr} \tag{12}$$

subject to capacity constraints in (8), (9), and the non-negativity of $D_{s\sigma}$ and $X_{sr}$.

We next note that the optimal solution will satisfy the constraint in (8) with equality. The reason is that for any stream $(s,\sigma)$ for which there is slack in (8), we may hold fixed all $X_{sr}$ for the stream, so that the total carried bandwidth on the stream is fixed, while decreasing $D_{s\sigma}$ to the point where the slack disappears, by increasing price $P_{s\sigma}$ appropriately. Hence, revenue from stream $(s,\sigma)$ is increased, without there being any impact on other streams.

By related reasoning, it can also be shown that the optimal solution will satisfy (9) with equality. This is a consequence of the assumption (see Section 2.2) that links are admissible routes for its end-points for at least one service. We argue that the demand and the carried bandwidth on the direct link may be increased equally by an amount that makes the unused capacity on the link vanish. The increase in demand is achieved by dropping the price, which, by virtue of the assumption that elasticity exceeds unity, leads to increased revenue.

With this observation, the joint pricing and routing problem becomes:

$$\max_{\{D_{s\sigma}\},\{X_{sr}\}} W = \sum_{s,\sigma} A_{s\sigma}^{1/\epsilon_{s\sigma}} D_{s\sigma}^{(\epsilon_{s\sigma}-1)/\epsilon_{s\sigma}} \tag{13}$$

subject to

$$\sum_{r\in\mathcal{R}(s,\sigma)} X_{sr} = D_{s\sigma} \qquad \forall(s,\sigma) \tag{14}$$

$$\sum_{s,\sigma}\sum_{r\in\mathcal{R}(s,\sigma):l\in r} X_{sr} = C_l \quad \forall l \tag{15}$$

$$D_{s\sigma} \geq 0 \qquad\qquad \forall(s,\sigma) \tag{16}$$

$$X_{sr} \geq 0 \qquad\qquad \forall(s,\sigma) \text{ and } \forall r\in\mathcal{R}(s,\sigma) \tag{17}$$

Note that as a consequence of $\epsilon_{s\sigma} > 1$, $\forall(s,\sigma)$, the objective function is a concave, monotonically increasing, differentiable function of $D_{s\sigma}$, and that the constraints are linear in the decision variables. The above is a special case of the class of "Concave Programming" problems, which have attractive properties; see for instance, [4, chapter 7]. Also, [16] shows that effective algorithms exist for concave programming.

### 3.2. Shadow Costs and Minimum Cost Routing

We follow Lagrange's Method [4], [7] in concave programming. Define the Lagrangian:

$$\begin{aligned} L(\boldsymbol{D},\boldsymbol{X},\boldsymbol{\mu},\boldsymbol{\lambda}) &= \sum_{s,\sigma} A_{s\sigma}^{1/\epsilon_{s\sigma}} D_{s\sigma}^{(\epsilon_{s\sigma}-1)/\epsilon_{s\sigma}} + \sum_{s,\sigma}\mu_{s\sigma}(\sum_{r\in\mathcal{R}(s,\sigma)} X_{sr} - D_{s\sigma}) \\ &+ \sum_{l}\lambda_l(C_l - \sum_{s,\sigma}\sum_{r\in\mathcal{R}(s,\sigma):l\in r} X_{sr}) \end{aligned} \tag{18}$$

Here $\{\mu_{s\sigma}\}$, $\{\lambda_l\}$ are the Lagrange multipliers associated respectively with the constraints implied by demand satisfaction in (14) and the link capacities in (15).

For the Kuhn-Tucker Theorem to hold, we require the constraint qualification, sometimes called the Slater condition, to hold. This requires that the matrix $\boldsymbol{G}$ is of maximum rank, where $\boldsymbol{G}$ is defined to be such that its application to the vector of decision variables ($\boldsymbol{D}$,$\boldsymbol{X}$) gives the left hand side of the constraints in (8) and (9). It can be shown that provided the admissible route sets satisfy simple non-degeneracy conditions, the constraint qualification is satisfied.

A main result from Concave Programming states that if ($\boldsymbol{D}$,$\boldsymbol{X}$) maximizes revenue $W$, then there exists ($\boldsymbol{\mu}$,$\boldsymbol{\lambda}$), such that

1. ($\boldsymbol{D}$,$\boldsymbol{X}$) maximizes $L$ in (18) without any constraints, i.e. for all streams $(s,\sigma)$

$$\frac{\partial L}{\partial D_{s\sigma}} \leq 0,\ D_{s\sigma} \geq 0 \quad \text{with complementary slackness,} \tag{19}$$

$$\frac{\partial L}{\partial X_{sr}} \leq 0,\ X_{sr} \geq 0 \quad \text{with complementary slackness,} \tag{20}$$

and

2. $\mu_{s\sigma} > 0 \quad \forall(s,\sigma),$ (21)

$$\lambda_l > 0 \quad \forall l \tag{22}$$

The Lagrange multipliers $\{\mu_{s\sigma}\}$ and $\{\lambda_l\}$ are shadow costs with properties that are discussed next. First, from (19) and (20) we have the following important relationship between the shadow cost $\mu_{s\sigma}$ and the optimal price $P_{s\sigma}$ for the stream $(s,\sigma)$:

$$P_{s\sigma} = (\frac{A_{s\sigma}}{D_{s\sigma}})^{1/\epsilon_{s\sigma}} = \frac{\epsilon_{s\sigma}}{\epsilon_{s\sigma}-1}\mu_{s\sigma} \quad \forall(s,\sigma) \tag{23}$$

Next, for each $(s,\sigma)$ and for each admissible route $r$ for which the (source, destination) is $\sigma$, we have:

$$\text{either} \quad X_{sr} > 0 \quad \text{and} \quad \sum_{l\in r} \lambda_l = \mu_{s\sigma} \tag{24}$$

$$\text{or} \quad X_{sr} = 0 \quad \text{and} \quad \sum_{l\in r} \lambda_l \geq \mu_{s\sigma} \tag{25}$$

The above condition is of considerable interest. To examine its consequences, begin by interpreting $\mu_{s\sigma}$ as the *minimum route cost* for $(s,\sigma)$. Next, we interpret $\lambda_l$ as the *link cost*. Finally, for any route $r$, we define the *route cost* to be the sum of link costs, i.e., $\sum_{l\in r} \lambda_l$. Thus, the condition in (24) states that if in the optimal solution an admissible route $r$ for the stream $(s,\sigma)$ is "active", in the sense that it carries traffic, then necessarily the route cost equals the minimum route cost for the stream. Thus, from the properties of the Concave Programming solution we have obtained link costs and a "minimum cost routing" policy, which is optimal. We will return to this policy later.

We end this section by emphasizing two points: first, the intimate relation between optimal stream prices, $\{P_{s\sigma}\}$, and the minimum route costs for streams $\{\mu_{s\sigma}\}$, and, second, link costs do not depend on the service type, which simplifies the implementation of the optimal routing policy.

## 4. OPTIMAL ROUTING

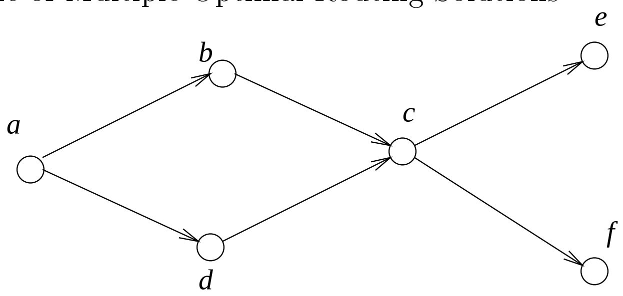
### 4.1. Procedures for Minimum Cost Routing

We develop further the minimum cost routing policy. In particular, we determine the amount of carried traffic on each of possibly several admissible, minimum cost routes, the occurrence of which is not uncommon.

We assume that an initial phase has been completed, and that it has yielded the optimal demands on streams, $\{D^*_{s\sigma}\}$, minimum route costs for streams $\{\mu_{s\sigma}\}$, and the link costs $\{\lambda_l\}$. In the second phase, we begin by defining the *minimum cost admissible route set* $M(s,\sigma)$: for all $(s,\sigma)$

$$M(s,\sigma) \equiv \{r | r \in \mathcal{R}(s,\sigma) \cap \sum_{l\in r} \lambda_l = \mu_{s\sigma}\} \tag{26}$$

Figure 1. An Example of Multiple Optimal Routing Solutions




$M(s, \sigma)$ is typically a considerably thinned subset of the set of admissible routes. Next, we solve the following systems of linear equations to obtain the carried load, $X_{sr}$, on each minimum cost route:

$$\sum_{r \in M(s,\sigma)} X_{sr} = D^*_{s\sigma} \qquad \forall (s, \sigma) \tag{27}$$

$$\sum_{s,\sigma} \sum_{r \in M(s,\sigma): l \in r} X_{sr} = C_l \qquad \forall l \tag{28}$$

Note that a solution to (27) and (28) is guaranteed. However, more than one solution may exist. We next consider an additional, optional phase to select a solution that is sometimes preferred for implementation purposes.

## 4.2. Procedures for Minimum Traffic Splitting

Given a set of optimal demands for streams, there may exist several different routing arrangements that yield the maximum of the objective function. As illustration, consider the example in Figure 1. Suppose in the optimal solution, $D_{(a,e),s} = D_{(a,f),s} = 1$, and the minimum cost admissible route sets for service $s$ are:

$$\begin{aligned} M(s, (a, e)) &= \{(a - b - c - e), (a - d - c - e)\}, \text{ and} \\ M(s, (a, f)) &= \{(a - b - c - f), (a - d - c - f)\} \end{aligned}$$

Then for any $x$ $(0 \leq x \leq 1)$,

$$\begin{aligned} X_{s,(a-b-c-e)} &= x, \quad X_{s,(a-d-c-e)} = 1 - x \\ X_{s,(a-b-c-f)} &= 1 - x, \; X_{s,(a-d-c-f)} = x \end{aligned}$$

is an optimal routing for $D_{(a,e),s}$ and $D_{(a,f),s}$. Each demand is carried on a single route if $x = 0$ or $x = 1$, and the demand is split between two admissible routes if $0 < x < 1$.

From one point of view, the implementation is more burdensome if demands are split, i.e., more than one route is used for a common (source, destination). The latency may vary with the routes, making it difficult to maintain quality of service of a stream, and requiring additional buffers to smooth the variations. Furthermore, some commonly used routing protocols, such as OSPF[15], only distribute traffic evenly among the candidate routes[5], and are not able to split a demand in arbitrary proportions in correspondence with the optimal solution.

Splitting the demand for a stream across multiple routes is generally unavoidable in the optimal solution. However, by solving the following optimization problem, we can always select the solution that gives the optimal revenue while minimizing the amount of split traffic.

$$z = \min_{\{X_{sr}\}} \sum_{s,\sigma} w_{s\sigma} \sum_{r \in M(s,\sigma)} (D^*_{s\sigma} - X_{sr}) X_{sr} \tag{29}$$

subject to:

$$\sum_{r \in M(s,\sigma)} X_{sr} = D^*_{s\sigma} \quad \forall (s,\sigma) \tag{30}$$

$$\sum_{s,\sigma} \sum_{r \in M(s,\sigma): l \in r} X_{sr} \le C_l \qquad \forall l \tag{31}$$

$$0 \le X_{sr} \le D^*_{s\sigma} \qquad \forall r \in M(s,\sigma) \tag{32}$$

where $D^*_{s\sigma}$ are optimal demands and $M(s,\sigma)$ are the minimum cost admissible route sets, both obtained from the preceding optimization phase. The penalty for splitting demand $D^*_{s\sigma}$ is the positive parameter $w_{s\sigma}$.

Notice that for the stream $(s,\sigma)$, $\sum_{r:r \in M(s,\sigma)} (D^*_{s\sigma} - X_{sr}) X_{sr}$ takes its minimum value of 0 if demand $D^*_{s\sigma}$ is carried on a single-path. Therefore, solving the problem in (29) eliminates any avoidable traffic-splitting (a traffic-splitting is considered to be avoidable if carrying that traffic on a single-path route will not induce a split of another otherwise non-split traffic). Notice that for different $\sigma$ and $s$, $w_{s\sigma}$ can take different values. In cases where it is less advisable to split traffic of one service (e.g. voice) than others (e.g. best-effort data), the weights may be used to attach a larger penalty.

## 5. CASE STUDIES

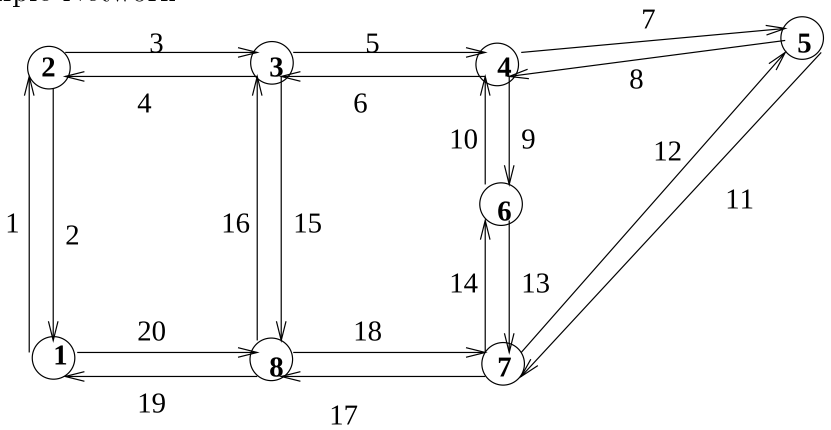
### 5.1. Assumptions, Parameters, and Design

In this section, we demonstrate properties of the optimal solution through numerical examples. We consider a sample network shown in Figure 2, which has eight nodes and 20 directed links. The node and link indices are indicated in the figure. The network offers two different services, say voice and data, indexed by $s = 1, 2$, respectively. Assume that the admissible route sets are as follows: the voice service can only be carried on minimum-hop routes, while the data service may be carried on routes with up to five hops.

In the constant price elasticity formulation of the demand function for each service(see Section 3.1), we make the assumption here that the demand potential $A_{s\sigma}$ may depend on both $s$ and $\sigma$, while price elasticity depends only on service type, i.e. $\epsilon_{s\sigma} = \epsilon_s$.

We construct a base case by letting $C_l = 400$ for all $l$, and $A_{1\sigma} = 2000$, and $A_{2\sigma} = 200$ for all $\sigma$, and $\epsilon_1 = 1.05$; $\epsilon_2 = 1.5$. We vary these parameters in the following different studies.

Figure 2. Sample Network




Specifically, we investigate the influences of price elasticity and uniform capacity expansion. By "uniform capacity expansion", we mean the scaling in which the capacities of all links grow in proportion to a common, increasing multiplier. Such expansions occur, for instance, when a carrier replaces existing systems with equipment of a later generation. For example, capacities of all network links will grow uniformly by a factor of 4 when 2.5Gbits/sec optical transmission systems are replaced by 10Gbits/sec systems. Now, recall from Section 2.1 that new data services are more elastic with respect to price changes than the traditional voice service. The combined effects of different price elasticities and uniform capacity expansion are important in multi-service networks. We examine their impacts on the optimal prices and demands, and on the optimal routing.

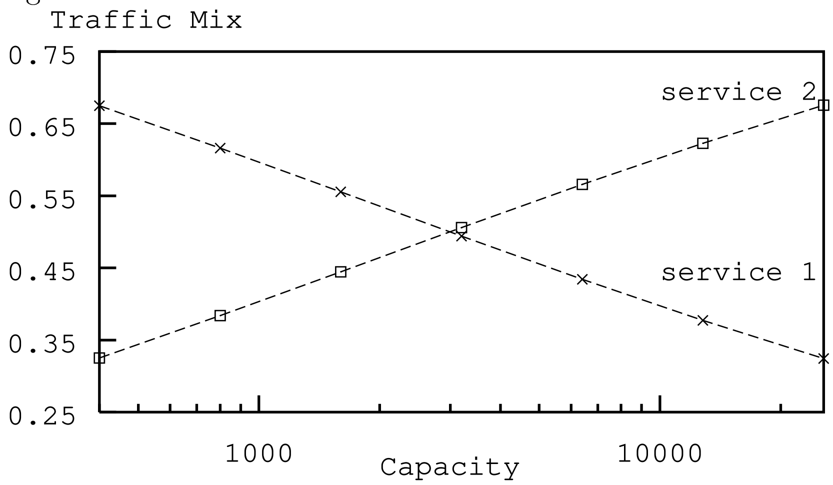
### 5.2. Effects on Traffic Mix

We first consider how uniform capacity expansion changes traffic composition. We let $C_l$ start at 400, and in each step we double it till the capacity of 25600 is reached. In Figure 3, the horizontal axis is capacity in log-scale.

For each service, we sum demands for all (source, destination) pairs, and divide the sum by the total amount of traffic in the network, i.e., $\dfrac{\sum_\sigma D_{s\sigma}}{\sum_t \sum_\sigma D_{t\sigma}}$, $s = 1, 2$. The two lines in the figure show the respective ratios for the two services. At the lowest capacity level, about 70% of total traffic is of service 1. The ratio declines when capacity increases, and is down to a value just above 30% when the capacity reaches 25600. In contrast, the percentage of service 2 traffic rises from about 30% to 70%.

The change of traffic mix is caused by the difference in price elasticity between the two services, given the relationship between prices and shadow costs in the optimal solution (see Section 3.2). As capacity increases, contraints (9) become less stringent, so shadow costs, which are Lagrange multipliers of those constraints, become smaller. Based on (23), prices will decline in the same proportion with route shadow costs, which are linear combinations of link shadow costs and common to all services (24). Nevertheless, from (5), with the same percentage decrease in price, service 2 (data), which has a higher price

Figure 3. Change of Traffic Mix




elasticity, will have a higher percentage increase in carried demand than service 1 (voice). Therefore, as we keep doubling link capacities, data streams gradually dominate voice streams in the total traffic bandwidth.

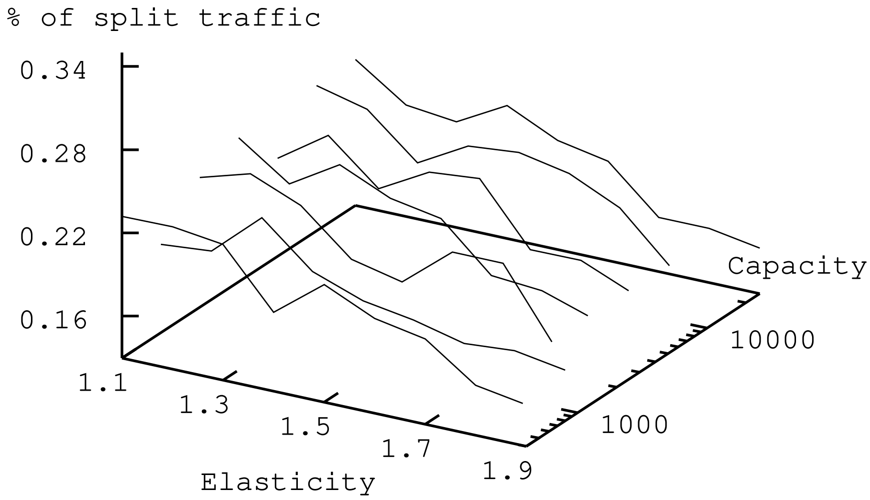
### 5.3. Effects on Route Splitting

In Section 4, we discussed techniques to reduce splitting traffic among alternative minimum-cost admissible routes. We now examine how uniform capacity expansion affects the minimum amount of split traffic.

The framework here is the same as in the base case, except that $A_{1\sigma} = 50$. The purpose of resetting $A_{1\sigma}$ is to create a starting case in which a significant portion of total traffic is carried on multiple routes. The ratio of split traffic, calculated by dividing the total amount of split traffic by the total traffic, is 21.44%. We then vary price elasticity from 1.1 to 1.9 in increments of 0.1. We also double network capacity uniformly from the base value of 400 to 25600. Figure 4 shows the split ratio at different capacity levels, given different values of price elasticity.

As shown in the figure, uniform capacity expansion has little impact on the route splitting. At each price elasticity level, there is no systematic decline in the percentage of split traffic, even though link capacities are increased by more than 50 times. The result suggests that the ability to distribute a stream, possibly non-uniformly, to multiple paths is an important traffic engineering function that will not be obsolesced by having "fat" bandwidth pipes in future. However, at each capacity level, the percentage of split traffic decreases (non-uniformly, however) as price elasticity increases. The explanation of this observation can be found in [14].

Figure 4. Percentage of Split Traffic




## 6. CONCLUSIONS

We have developed a revenue-maximization model for jointly optimizing pricing and routing in a multi-service network. For this model, we have obtained the optimal solution, and derived its salient properties. We have defined and derived link shadow costs as common bases for unifying pricing and routing decisions. We have shown in a numerical example that the service with the highest price-elasticity, such as data, will dominate total traffic when network capacity is scaled to increase uniformly. We have shown in another numerical example that the significance of traffic splitting does not diminish in the presence of uniform capacity expansion.

Our joint optimization model can be connected to other decision problems related to providing network services. For example, the use of shadow costs, derived from this model, to regulate capacity provisioning is being considered.

## REFERENCES


1. Ahuja, R. K., Magnanti, T. L., and J. B. Orlin, *Network Flows.* Prentice-Hall, Inc., Upper Saddle River, NJ, 1993.
2. Aldebert, M., Ivaldi, M., and C. Roucolle, Telecommunication Demand and Pricing Structure, *Proceedings of the 7th International Conference on Telecommunications*

*Systems: Modeling and Analysis.* pp. 254-267, Nashville, TN, March 1999.

3. Courcoubetis, C. A., Dimakis, A., and M. I. Reiman, Providing Bandwidth Guarantees Over a Best-effort Network: Call Admission and Pricing, *Proceedings of IEEE INFOCOM2001*, pp. 459-467, April 2001.
4. Dixit, A. K., *Optimization in Economics Theory.* 2nd edition, Oxford University Press, Inc., New York, 1990.
5. Fortz, B. and M. Thorup, Internet Traffic Engineering by Optimizing OSPF Weights, *Proceedings of IEEE INFOCOM2000.* pp. 519-528, 2000.
6. Greene W., *Econometric Analysis.* 2nd edition, Collier-Macmillan, London, 1999
7. Intriligator, M. D., *Mathematical Optimization and Economic Theory.* Englewood Cliffs, Prentice-Hall, NJ, 1971.
8. Kelly, F., Maulloo, A., and D. Tan, "Rate Control in Communication Networks: Shadow Prices, Proportional Fairness and Stability," *Journal of the Operational Research Society.* 49, pp. 237-252, 1998
9. Lagoa, C. and H. Che, "Decentralized Optimal Traffic Engineering in the Internet," *Computer Communications Review.* vol. 30, No. 5, pp. 39-47, October, 2000.
10. Lanning, S. G., Mitra, D., Wang, Q., and M. H. Wright, "Optimal Planning for Optical Transport Networks", *Phil.Trans.Royal Soc.London A.* Vol.358, No.1773, pp. 2183-2196, August 2000.
11. Lanning, S. G., Neuman, W. R., and S. O'Donnell, "A Taxonomy of Communications Demand," *Proceedings of 27th Telecommunications Policy Research Conference.* Alexandria, VA, September 1999.
12. MacKie-Mason, J. K. and H. Varian, "Pricing Congestible Resources," *IEEE Journal of Selected Areas in Communications.* vol. 13, no. 7, pp. 1141-1149, September 1995.
13. Mitra, D. and K. G. Ramakrishnan, "A Case Study of Multiservice Multipriority Traffic Engineering Design for Data Networks", *Proceedings of IEEE GLOBECOM 99.*, pp. 1077-1083, December 1999.
14. Mitra, D., Ramakrishnan, K. G., and Q. Wang, "Combined Economic Modeling and Traffic Engineering: Joint Optimization of Pricing and Routing in Multi-Service Networks", *Bell Labs Technical Memorandum*, 2001.
15. Moy, J. T., *OSPF Anatomy of an Internet Routing Protocol.* Addison-Wesley, Reading, MA, 1998.
16. Vavasis, A., *Nonlinear Optimization: Complexity Issues.* Oxford University Press, New York, 1991.
17. Wang, Q., Peha, J. M., and M. A. Sirbu, "Optimal Pricing for Integrated Services Networks," *Internet Economics.* edited by L. W. McKnight and J. P. Bailey, pp. 352-376, MIT Press, Cambridge, MA, 1997.