On Optimal Sizing of Tiered Network Services

Qian Lv and George N. Rouskas
Department of Computer Science, North Carolina State University, Raleigh, NC 27695-8206, USA
Email: {qlv,rouskas}@ncsu.edu

***Abstract*— We develop an economic model for networks offering tiered services and we formulate the problem of selecting the service tiers from three perspectives: one that considers the users' interests only, one that considers only the service provider's interests, and one that considers both simultaneously, i.e., the interests of society as a whole. We also present dynamic programming algorithms that solve these problems optimally. Our work provides a theoretical framework for reasoning about Internet tiered services, as well as a practical toolset for network providers to develop customized menus of service offerings.**

I. Introduction

Internet service providers have introduced several forms of a *tiered service*, in which users may select from a small set of service levels (*tiers*) which offer progressively higher bandwidth with a corresponding increase in price. The introduction of tiered service has important engineering (e.g., in terms of operation, control, and management) and financial implications for the network provider. A continuous-rate network must be designed to accommodate any arbitrary rate of service requested by the users. In a tiered-service network, on the other hand, a wide range of core functions, including equipment configuration, traffic engineering, quality of service (QoS) support and service level agreements, billing, and customer support are simplified, enabling the providers to scale their operations to millions of customers. Currently, service tiers are either based on the bandwidth hierarchy of the underlying network infrastructure (e.g., DS-1, OC-3, etc.), or are determined in some *ad-hoc* manner (e.g., the ADSL tiers available from various providers). In [5] we developed a systematic framework for tiered-service networks by adopting the network operator's perspective. We have also demonstrated the benefits of tiered service for traffic engineering [6] and packet scheduling for QoS [7].

In this paper, we extend our work in [5] by developing an economic model for tiered-service networks that allows us to formulate the problem of selecting the service tiers from three perspectives: one that considers the users' interests only, one that considers only the service provider's interests, and one that considers both simultaneously, i.e., the interests of society as a whole. We also present dynamic programming algorithms that solve these problems optimally. Our work provides a theoretical framework for reasoning about Internet tiered services, as well as a practical toolset for network providers to develop customized menus of service offerings that cater to user needs while ensuring that both parties are satisfied.

This work was supported by the NSF under grant CNS-0434975.

The rest of the paper is organized as follows. In Section II, we introduce the tiered-service network we consider in this study. In Section III, we introduce an economic model for tiered-service networks that takes into account the user's perspective, the provider's perspective, or both. We also formulate and solve three corresponding problems for selecting the set of service tiers optimally. We present numerical results in Section IV and we conclude the paper in Section V.

II. The Tiered-Service Network

We consider a network with N users. The network provides a service characterized by the amount of bandwidth x allocated to each user, as is typical of current residential (e.g., DSL or cable modem) and business Internet access services (e.g., T1, T3, or higher). We assume that users may request any amount of bandwidth depending on their needs and their willingness or ability to pay the corresponding service fee. We let d_i denote the bandwidth request (i.e., *demand*) of user i, and define the demand vector $D =< d_1, d_2, \cdots, d_N >$, where we have labeled the user demands in non-decreasing order of requested bandwidth, $d_1 \leq d_2 \leq \cdots \leq d_N$. In this paper we assume that the demand vector is known to the service provider.

The network provider offers K bandwidth levels (tiers) of service, where typically K is a small integer such that $K \ll N$. We define $S =< s_1, s_2, \cdots, s_K >$ as the vector of service tiers offered by the network provider; without loss of generality, we assume that the service tiers are labeled such that $s_1 < s_2 < \cdots < s_K$. For notational convenience, we also define the "null" service tier $s_0 = 0$.

With tiered service, a user i with bandwidth demand d_i will have to subscribe to service tier s_j such that $s_{j-1} < d_i \leq s_j$ so as to experience a QoS that meets or exceeds its requirements. Figure 1 shows a sample mapping from a vector of 13 bandwidth demands to a vector of 6 service tiers. Note that the network provider needs to provide each user i with additional bandwidth $(s_j - d_i) \geq 0$, and will typically incur higher costs for doing so; consequently, the provider will be inclined to select the service tiers so as to recoup these costs (and make a profit). On the other hand, user i subscribes to a service (i.e., s_j) that is at least as good as the one requested (i.e., d_i), but the additional value, if any, that the user receives may be offset by the higher cost of the service. Our aim is to apply economic theory to capture analytically these tradeoffs, and to develop techniques to select the service tiers in a manner that accounts for both the users' and providers' perspectives.

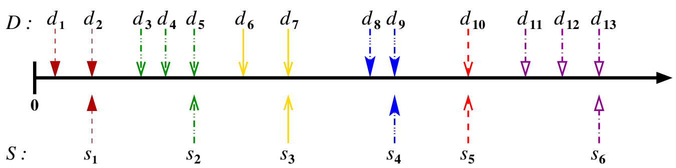

Fig. 1. Sample mapping of bandwidth demands to service tiers

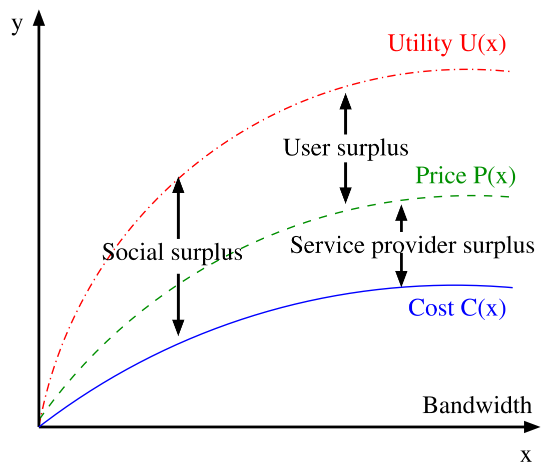

Fig. 2. Utility, cost, and price functions

To develop an economic model for tiered-service networks, we assume the existence of three non-decreasing functions of bandwidth x, as shown in Figure 2. The *utility* function, $U(x)$, is a measure of the value that users receive from the service, and it stands for their willingness to pay for the service. The *cost* function, $C(x)$, represents the bandwidth cost incurred by the provider for offering the service. Finally, the *price* function, $P(x)$, represents the amount that the service provider charges for the service. Figure 2 shows that $U(x)$ lies above $P(x)$ (otherwise users would not be willing to pay for the service), and in turn $P(x)$ lies above $C(x)$ (otherwise providers would not be inclined to offer the service); however, our results are obtained for general functions $U(x)$, $P(x)$, and $C(x)$, independent of the relative behavior of the corresponding curves. We make the reasonable assumption that utility, cost, and price are all expressed in the same units (e.g., US$). Note that utility and cost typically depend only on the user and service provider, respectively, but that price is the result of market dynamics and the relative bargaining power of users and service providers.

III. Economic Model for Tiered-Service Networks

We now use concepts from economics to describe the relationship between users and service providers, and we propose a series of optimization problems for determining the bandwidth tiers in tiered-service networks. We also illustrate how to solve these problems using dynamic programming to obtain a set of optimal bandwidth tiers.

Consider now the demand-supply relationship between the users and network service providers. On the one hand, users want to maximize the utility they obtain from the service while keeping the fee they have to pay to the service provider as low as possible; in economic terms, users want to maximize the *user surplus* [2], [3], defined as the difference between the utility they obtain from the service and the price they have to pay for it. On the other hand, the network providers' objective is to charge a high fee so as to offset the cost of offering the service and make a profit; in other words, service providers want to maximize the *service provider surplus* [2], [3], defined as the difference between price and cost. The concepts of user surplus and service provider surplus are illustrated in Figure 2.

From the point of view of the society as a whole, it is preferable to maximize the overall *social welfare*, defined as the sum of the user surplus plus the provider surplus (see also Figure 2). We will refer to the social welfare as *social surplus* [2], [3]. Once the maximum social surplus has been determined, the users and service providers may negotiate its division into user and service provider surpluses through bargaining.

In the tiered-service network under consideration, the problem of maximizing the surplus of users, service providers, or society, amounts to selecting appropriately the set of service tiers to be offered. The following three subsections formulate each of these optimization problems and present efficient algorithms for solving them optimally.

A. Service Tier Optimization: The User Perspective

Let us first consider the problem of optimally selecting the service tiers from the users' point of view. We make the assumption that the utility function $U(x)$ and the price function $P(x)$ are known and fixed; we will address the issue of determining an optimal price function shortly. Based on our earlier discussion, the objective of each network user is to maximize its surplus. Considering all the users in the network as a whole, the objective is to select the set of service levels so as to maximize the aggregate surplus, i.e., the sum of the individual user surpluses. This optimization problem, which we will refer to as the *Service Tier Optimization for Users (STO-U)* problem, can be formally expressed as follows.

Problem 3.1 (STO-U): Given a vector D of N bandwidth demands, $d_1 \leq \cdots \leq d_N$, an integer number $K < N$ of service tiers, a utility function $U(x)$, and a price function $P(x)$, find a service tier vector $S =< s_1, \cdots, s_K >$ that maximizes the objective function (aggregate user surplus):

$$F_U(S) \quad = \quad \sum_{j=1}^{K} \mid D_j \mid (U(s_j) - P(s_j)) \qquad (1)$$

subject to the constraints:

$$s_{j-1} \quad < \quad d_i \quad \leq \quad s_j, \quad d_i \in D_j, \ j = 1, \cdots, K \qquad (2)$$

where D_j is the set of demands mapped to service tier $s_j, j = 1, \cdots, K$, and $s_0 = 0$ is the "null" service tier.

The STO-U problem is an instance of the directional k-median problem we introduced in [5], and can be solved using a dynamic programming algorithm similar to the one we presented there. For completeness, we describe the dynamic programming formulation next.

Define $\Phi(n,k)$ as the maximum value of the objective function (1) when the number of users (demands) is n and the number of service tiers is $k \le n$. Then, it is possible to solve STO-U by using the following dynamic programming formulation to compute $\Phi(n,k)$ recursively:

$$\Phi(n,1) = n(U(d_n) - P(d_n)), \quad n = 1,\cdots,N \tag{3}$$

$$\Phi(1,k) = U(d_1) - P(d_1), \quad k = 1,\cdots,K \tag{4}$$

$$\Phi(n,k+1) = \max_{q=k,\cdots,n-1} \{\Phi(q,k) + (n-q)(U(d_n) - P(d_n))\}$$
$$k = 1,\cdots,K-1;\ n = 2,\cdots,N \tag{5}$$

Expression (3) states that when $K = 1$, due to constraints (2), the optimal service tier is equal to the largest demand d_n; in this case, all n users obtain utility $U(d_n)$ and pay price $P(d_n)$. Expression (4) states that if there is only one user with demand d_1, the user will select a service tier equal to d_i[1]. The recursive equation (5) can be explained by noting that the $(k+1)$-th service tier must be equal to the largest demand d_n. If the k-th service tier is equal to d_q, $q = k,\cdots,n-1$, the aggregate user surplus is given by the expression in brackets in the right-hand side of (5), since $n-q$ demands are mapped to service tier d_q. Taking the maximum over all values of q provides the optimal value.

A straightforward implementation of the recursion (3)-(5) takes time $O(KN^2)$. By exploiting the fact that the $N \times K$ matrix Φ satisfies the concave Monge condition [1], we were able to develop an $O(KN)$ implementation of the dynamic programming algorithm that is efficient when the number N of users is large; for the details of the implementation, the reader is referred to [5].

B. Service Tier Optimization: The Provider Perspective

The goal of the service provider is to maximize its aggregate surplus, $P(x)-C(x)$, over all N users. Given a price function $P(x)$ and a cost function $C(x)$, the problem of determining an optimal vector of K service tiers can then be formulated in a manner similar to expressions (3)-(5) and solved using a similar dynamic programming algorithm.

We now develop a more realistic formulation of the problem, based on the observation that, in general, the total cost to the network provider of offering K tiers of service consists of two components. The first component is due to the cost of the bandwidth: the more bandwidth acquired by the users, the higher the cost. We use the nondecreasing function $C(x)$ to denote this cost, representing the link cost for carrying user traffic, as well as the switching cost at routers. The second component captures the cost of software and hardware mechanisms at the routers for supporting a given number K of service tiers. Specifically, we assume that the *incremental* cost (e.g., due to the additional queueing structures, policing mechanisms, control plane support, etc.) of offering one additional service tier is equal to α. Hence, the total cost for K tiers is[2] αK.

Based on the above discussion, the provider's optimization problem, which we refer to as the *Service Tier Optimization for Providers (STO-P)* problem, is defined as follows.

Problem 3.2 (STO-P): Given a vector D of N bandwidth demands, $d_1 \le \cdots \le d_N$, a price function $P(x)$, a bandwidth cost function $C(x)$, and a per-service tier cost α, find the number $K, K \le N$, of service tiers and an optimal service vector $S =< s_1, s_2, \cdots, s_K >$ that maximize the objective function (aggregate provider surplus):

$$F_P(K,S) = \left\{ \sum_{j=1}^{K} \mid D_j \mid (P(s_j) - C(s_j)) \right\} - \alpha K \tag{6}$$

subject to the constraints:

$$s_{j-1} < d_i \le s_j, \quad d_i \in D_j,\ j = 1,\cdots,K \tag{7}$$

where D_j is again the set of demands mapped to service tier $s_j, j = 1,\cdots,K$.

The objective function (6) is derived under the assumption that the bandwidth cost to the network provider is simply the sum, over all users, of the individual cost $C(x)$ of providing service x to a user in isolation. An alternative objective function can be obtained by assuming that the bandwidth cost to the provider is just the cost of providing the aggregate bandwidth over all users:

$$F'_P(K,S) = \sum_{j=1}^{K} |D_j| P(s_j) - C\left(\sum_{j=1}^{K} |D_j| s_j\right) - \alpha K \tag{8}$$

The STO-P optimization problem is a generalization of the directional k-median problem [5] as the number K of service tiers is now a variable. STO-P belongs to the family of uncapacitated facility location problems [4]. This family of problems include as part of the input a facility-build cost (in our case, a service tier cost α), and the objective is to optimize jointly the number of facilities and their locations. Of course, due to the directionality constraint (i.e., the requirement that a user subscribe to a service tier at least as large as its bandwidth demand), algorithms for existing location problems in which a demand may be served by any facility, to the left or right of it in the real line, cannot be applied to the STO-P problem.

We now present a solution approach for the STO-P problem with the objective function (6) that is based on dynamic programming; the STO-P problem with the alternative objective

[1] Strictly speaking, expression (4) will optimize the objective function (1) only when the gap $U(x) - P(x)$ between the utility and price functions does not increase for $x > d_1$. In this case, whenever the number of service tiers is greater than or equal to the number of users, each user receives exactly the amount of service it requests. Since for most typical scenarios the utility of a service or product tends to increase slower than price (otherwise, users would purchase the largest amount of product or service offered), it is reasonable to use expression (4) in the dynamic programming solution. More generally, let $x^\star > d_1$ be a value such that $U(x^\star) - P(x^\star)$ is maximum over all $x > d_1$; then, one would set $\Phi(1,k)$ in (4) to $U(x^\star) - P(x^\star)$. We have not considered this alternative here.

[2] We emphasize that our formulations and results can be extended in a straightforward manner to any non-decreasing function $g(K)$ in place of αK.

function (8) can be solved in a similar manner. As in the previous subsection, we define $\Phi(n,k)$ as the maximum value of the aggregate service provider surplus in (6) when the number of users is n and the number of service tiers is $k \le n$. We can obtain $\Phi(n,k)$ using the following recursion:

$$\Phi(n,1) \;=\; n(P(d_n)-C(d_n))-\alpha, \quad n=1,\cdots,N \quad (9)$$

$$\Phi(1,k) \;=\; P(d_1)-C(d_1)-\alpha k, \quad k=1,\cdots,N \quad (10)$$

$$\Phi(n,k{+}1) = \max_{q=k,\cdots,n-1}\{\Phi(q,k){+}(n{-}q)(P(d_n){-}C(d_n))\}{-}\alpha$$
$$k=1,\cdots,N-1;\; n=2,\cdots,N \quad (11)$$

Expressions (9)-(11) are similar to (3)-(5), respectively. The main difference is the introduction of the service tier ("facility") cost α, which decreases the service provider surplus accordingly. For instance, in expression (11), the cost of the additional (i.e., $(k+1)$-th) service tier is accounted for in the right hand side by subtracting the value of parameter α. We also note that similar comments to the ones we made for expression (4) apply to expression (10) as well.

At the end of the recursion, the entries of the last row of the table Φ, i.e., the values of $\Phi(N,k), k=1,\cdots,N$, correspond to the optimal servicer provider surplus for the given demand vector D when there are k service tiers. Let $k^\star$ be the optimal value of k, i.e., a value such that $\Phi(N,k^\star) \ge \Phi(N,k)$ for all $k, k=1,\cdots,N$. The value $k^\star$ and the corresponding service tiers comprise the optimal solution to the STO-P problem.

Since this $N \times N$ matrix Φ also satisfies the concave Monge condition [1], the time complexity of the dynamic programming algorithm is $O(N^2)$ using the implementation we described in [5]. Finding the optimal value $k^\star$ by searching the last row of matrix Φ takes time $O(N)$, hence the overall time complexity of the algorithm is $O(N^2)$.

C. Service Tier Optimization: The Society Perspective

So far we have assumed that users and service providers may select the service tiers optimally based only on their own interests. In reality, this assumption may not be reasonable or practical. An optimal service vector for the users may not be acceptable to the service provider, and vice versa. Therefore, it is important to obtain a jointly optimal solution that takes into account the perspectives of both users and service providers. Furthermore, the optimization problems STO-U and STO-P take the price function $P(x)$ as input. In general, the price function is the result of negotiation between users and service providers, hence it may not be known in advance. We now show that considering the welfare of the society (i.e., users and providers) as a whole overcomes these difficulties, allows us to determine the optimal service tier vector without knowledge of the price function, and leads to an elegant approach for determining optimal prices for the service tier vector.

From the society's perspective, it is desirable to maximize the social surplus, i.e., the sum of user and service provider surpluses. Let us assume that the utility function $U(x)$ and the cost functions (i.e., bandwidth cost $C(x)$ and per-tier cost α) are known. Then, maximizing the social surplus leads to the following optimization problem, which we call the *Service Tier Optimization for Society (STO-S)* problem:

Problem 3.3 (STO-S): Given a vector D of N bandwidth demands, $d_1 \le \cdots \le d_N$, a utility function $U(x)$, a bandwidth cost function $C(x)$, and a per-service tier cost α, find the number $K, K \le N$, of service tiers and an optimal service vector $S =< s_1, s_2, \cdots, s_K >$ that maximize the objective function (social surplus):

$$F_S(K,S) \;=\; \left\{ \sum_{j=1}^{K} \mid D_j \mid (U(s_j)-C(s_j)) \right\} - \alpha K \quad (12)$$

subject to the constraints:

$$s_{j-1} \;<\; d_i \;\le\; s_j, \quad d_i \in D_j,\; j=1,\cdots,K \qquad (13)$$

where D_j is the set of demands mapped to service tier s_j.

This problem is identical to the STO-P problem, except that in the objective function (12) $U(x)$ is used whenever $P(x)$ is used in (6). Hence, STO-S can be solved in time $O(N^2)$ with the dynamic programming algorithm in (9)-(11). After solving the STO-S problem, we obtain an optimal service vector $S^\star =< s_1, s_2, \cdots, s_K >$ that maximizes the social surplus and depends only on the utility and cost functions provided by the users and network provider, respectively.

IV. Numerical Results

To illustrate our methodology for sizing of tiered services, we consider the market for broadband Internet access. Specifically, we assume that user demands are in the range [256 Kb/s, 6.1 Mb/s], typical of current broadband speeds in the United States. We consider two distributions for user demands, as in Figure 3: a uniform distribution, under which a user is equally likely to request any amount of bandwidth in the specified range, and a six-modal distribution in which user demands are concentrated around multiples of 1 Mb/s. In particular, the six-modal distribution is such that with probability 0.1167 a user demand will be in the range $[k{-}0.1, k{+}0.1], k=1,\cdots,6$ Mb/s, and with probability 0.3 a user demand will take any other value. We let the number of users $N=1000$.

We let the bandwidth cost function $C(x)=\mu x$ and the tier cost function $C(K)=\alpha K$. We let the utility function $U(x)=\lambda x^\gamma \log(x)$. Recall that utility stands for the users' willingness to pay, and in most cases it is an increasing, strictly concave, and continuously differentiable function of bandwidth. This function, which can be easily shown to have all three properties, was considered within the context of elastic traffic in [8]. The parameters λ and γ can be used to control the slope of the utility function $U(x)$; we use $\mu=0.5$, $\alpha=250$, $\lambda=10$, and $\gamma=0.4, 0.5$, and we discuss these choices shortly.

Due to space constraints, we only consider the STO-S problem. Let us consider the impact of the number K of tiers on the value of the social surplus. Figure 4 plots the

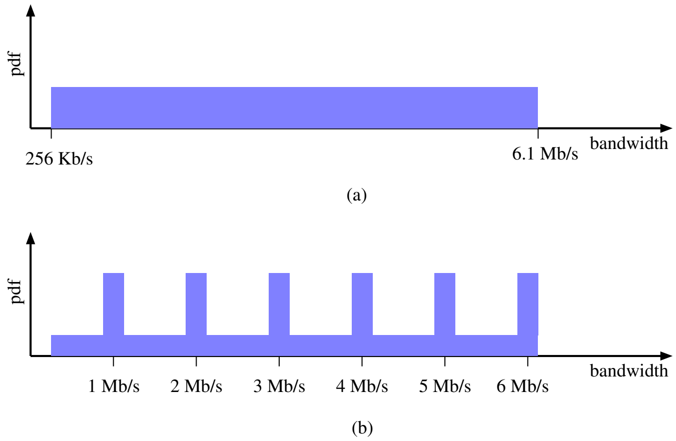

Fig. 3. User demand distributions: (a) uniform, (b) six-modal

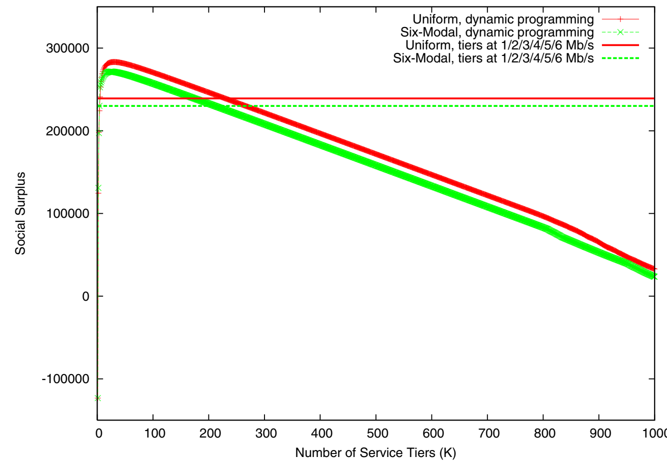

Fig. 4. Social surplus against K, $\gamma = 0.5$

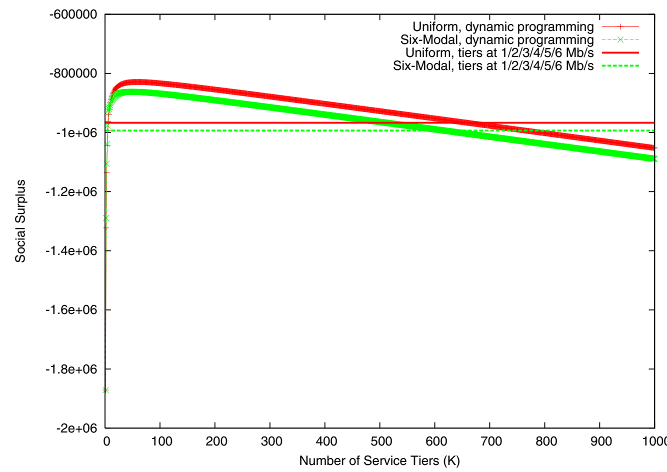

Fig. 5. Social surplus against K, $\gamma = 0.4$

social surplus obtained by the dynamic programming algorithm against K, for the utility function with $\gamma = 0.5$. Two curves are shown, one for demands drawn from the uniform distribution and one for demands drawn from the six-modal distribution of Figure 3. We observe that the social surplus initially increases with K, reaches a maximum value, and then starts to decrease. This behavior can be explained by noting that when K is small, the bandwidth cost $C(x)$ dominates the tier cost αK; therefore, introducing additional tiers allows the provider to better match its service offerings to the user demands, increasing the overall surplus. However, after K crosses a threshold (that depends on the values of parameters α, γ, λ, and μ), the tier cost starts to dominate, decreasing the provider's surplus and more than compensating for any increase in the user surplus. The behavior is similar for the two distributions, and for others not shown here.

Figure 4 also plots (as straight lines) the value of the social surplus for a simple service offering with six tiers at multiples of 1 Mb/s, for both demand distributions. As we can see, the social surplus achieved by this set of tiers is substantially lower than the maximum surplus determined by the dynamic programming algorithm; this observation is true even for the six-modal distribution of user demands. This example illustrates that selecting the service tiers using informal, *ad-hoc* approaches is likely to lead to suboptimal solutions; our methodology, on the other hand, is designed to find solutions that maximize the overall benefit to society.

Figure 5 is similar to Figure 4, except for the fact that we used $\gamma = 0.4$ in the utility function $U(x)$. For this value of γ, the utility function is lower than the cost function $C(x)$ over most of the range of user demands. Consequently, the social surplus is negative in this case. Nevertheless, our methodology remains valid, and the dynamic programming algorithm can be used to determine the service tiers that maximize the social surplus; this optimal set of tiers again outperforms the set of six tiers at multiples of 1 Mb/s.

V. Concluding Remarks

We proposed an economic model for tiered-service networks and developed dynamic programming algorithms to select the service tiers. Our approach provides insight into the selection of Internet tiered services, as well as a theoretical framework of practical importance to network providers.

References

[1] A. Aggarwal and J. Park. Notes on searching in multidimensional monotone arrays. In *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 497–512, 1988.

[2] R. Bade and M. Parkin. *Foundations of Microeconomics*. Addison-Wesley, 1st edition, 2001.

[3] B. C. Eaton, D. F. Eaton, and D. W. Allen. *Microeconomics*. Prentice Hall, 5th edition, 2002.

[4] G. P. Cornuejols G. L. Nemhauser and L. A. Wolsey. The uncapacitated facility location problem. In *P. Mirchandani and R. Francis (Editors), Discrete Location Theory*, pages 119–171. Wiley, 1990.

[5] G. N. Rouskas and N. Baradwaj. A framework for tiered service in MPLS networks. *Proc. IEEE INFOCOM 2007*, pages 1577–1585, May 2007.

[6] G. N. Rouskas and N. Baradwaj. TDM emulation in packet-switched networks. In *Proceedings of IEEE ICC*, pages 1911–1916, June 2007.

[7] G. N. Rouskas and Z. Dwekat. A practical and efficient implementation of WF^2Q+. In *Proceedings of IEEE ICC*, pages 172–176, June 2007.

[8] S. Shenker. Fundamental design issues for the future internet. *IEEE J. Selected Areas in Communications*, 13(7):1176-1188, Sep. 1995.