ABSTRACT

QIAN, LV. Economic Models for Tiered Network Services. (Under the direction of Dr. George N. Rouskas).

Tiered network services have been prevalent in current networking industry. The term *tiered service* means that the network operator only provides a small set of tiers (levels) - which offer progressively higher levels of service - to the customers each of whom will be mapped to one of the given tiers. In this thesis we focus on the economic issues for tiered network service. We first formulate the problem of selecting 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. We consider the solution to this problem under two cases: 1) the discrete case, i.e., each user demand is known to the service provider; 2) the stochastic case, i.e., the service provider only knows the probability distribution of the user demands. For both cases, we present accurate and efficient algorithms based on dynamic programming. After finding the set of (near-) optimal service tiers, we then employ game-theoretic techniques to find an optimal price for each service tier that strikes a balance between the conflicting objectives of users and service provider. This work provides a theoretical framework for reasoning about and pricing Internet tiered services, as well as a practical toolset for network providers to develop customized menus of service offerings.

We further consider some advanced economic topics in tiered network service. We notice that some network services may tend to be elastic, i.e., the users may value a given service differently and show different willingness to pay for the service. In this thesis, we assume that users are partitioned into some distinct classes. We develop an optimal algorithm to select jointly the set of service tiers and their prices so as to maximize the provider profits. Our research shows that introducing multiple tiers can be an effective market segmentation strategy that may lead to an increase in profits.

Another advanced topic in tiered network service is service bundling, which means the network providers combine several services together and sell them as a single package at a lower price than that if the services are sold separately. Based on tools from microeconomics and utility theory, we developed an efficient method to find tiered structures for bundles of network services with the objective of maximizing provider profits under user constraints.

Economic Models for Tiered Network Services

by

Qian Lv

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fullfillment of the
requirements for the Degree of
Doctor of Philosophy

Computer Science

Raleigh, North Carolina

2010

Approved By:

Dr. Khaled Harfoush	Dr. Wenye Wang
Dr. George N. Rouskas Chair of Advisory Committee	Dr. Rudra Dutta

DEDICATION

To my parents, wife, and daughter.

Biography

Qian Lv was born and grew up in Yueyang, Hunan, China. He received his B.S. degree and M.S. degree both on Electrical Engineering from China Institute of Metrology and Huazhong University of Science and Technology in 1999 and 2002, respectively. In 2005, he came to the department of Computer Science, North Carolina State University, to pursue his PhD degree.

He worked as a software engineer in Fiberhome Telecommunication Technologies in China, during 2002 to 2003, and interned in ABB and Sierra Wireless in USA in 2008 and 2009, respectively. During 2003 to 2005, he was a graduate research assistant in Nanyang Technological University, Singapore. His current research areas include Internet Economics and facility location analysis. He is a student member of ACM(Association of Computing Machine), IEEE(Institute of Electrical and Electronics Engineers) and SIAM(Society for Industrial and Applied Mathematics). He plans to be a software developer in IT industry after graduation.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. George N. Rouskas, for teaching me how to do research. His patience, encouragement, and in-depth knowledge has provided invaluable help for my research work. This thesis would have not been possible without his help and support. Besides my advisors, I would like to thank the rest of my thesis committee: Dr. Rudra Dutta, Dr. Khaled Harfoush, Dr. Wenye Wang, for their insightful comments and questions on my PhD work.

Much gratitude to my parents, Xinbao Lv and Guizhang Yang, who have been offered me support without any reservation, ever since I was born. To my wife, Haifeng Yu, for her help these years. To my daughter, Nicole Lv, whose born brought me too much happiness. To my sister, Jing Lv, brother-in-law, Yisong Wang, and nephew, Yuxiang Wang, for their long-time support and encouragement. To my uncle, Mingfang Lv, who has been standing by me in my every step.

I also want to thanks my friends in the department, including Mohan Iyer, Claris Castillo, Anjing Wang and Huan Luo, for their friendship. And thanks all the people who helped me during my PhD study.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

1.1 Background

Traditionally, Internet and Asynchronous Transfer Mode (ATM) network are designed to be continuous-rate, that is, they allocate bandwidth at very fine granularities such that the user can request any rate of service. The continuous-rate design offers the customers maximum flexibility in utilizing the available network capacity. For example, one customer may request a rate of 99.8Mbps of bandwidth, while the other may ask for 99.9Mbps - only 0.1Mbps gap - and the difference can be as small as 1 bps in the extreme case. Such requirement that supports bandwidth at such fine granularity, however, is a disaster for the network operators, because it is extremely difficult, or even impossible, to differentiate between the two users with very close bandwidth request. As a result, some of the network's key functions - such as robust traffic policing or accurate customer billing - are seriously undermined in case of a large, heterogeneous user population.

To address such challenges, most network operators developed tiered service by which the customers can only select from a small set of service tiers (i.e., levels) which offer progressively higher levels of service with a corresponding increase in price. With tiered service, the users with close requests will be mapped to the same tier with the next higher available rate. For example, the two customers requesting 99.8Mbps and 99.9Mbps will be both mapped to the next higher available rate, say, 100Mbps. Thus, the core function of the network, such as network management and equipment configuration, traffic engineering, service level agreements, billing and customer support, can be greatly simplified.

Currently, there are two kinds of tiered structure chosen by the major ISPs. Some ISPs use bandwidth hierarchy of the underlying transport network infrastructure, such as DS-1, DS-3, OC-3, etc. Other ISPs employ exponential tiering structures in which each tier offers twice the bandwidth of the previous one. One typical example of the exponential service structure is the various ADSL tiers - 384Kbps, 768Kbps, 1.5Mbps, 3Mbps, 6Mbps, etc - which are chosen by some major ISPs in US.

Note that tiered structures may be also be employed when the offered service is characterized by the traffic volume instead of bandwidth. For instance, Time-Warner launched a pilot program [12, 13, 30] in Beaumont, Texas in spring 2008, under which it charges customers based on the amount of traffic they transfer (including upload and download). Time Warner uses exponential tier structure for the pilot program, with the tiers placed in 5 GB, 10 GB, 20 GB, and 40 GB of monthly traffic.

The deployment of tiered service inevitably brings some economic issues, because the chosen of service tiers and the corresponding prices would affect the interests for both the customers and service provider. From the customer's view, they prefer lower price and more comfortable service. From the service provider's view, they would like to spend less money on network devices and charge higher price on the service. The bandwidth hierarchy tier structure or exponential tier structure may be far from the optimal solution for either the customers or service provider. The bandwidth hierarchy is a natural arrangement for the service provider; it is unlikely that the hierarchical rates designed decades ago for voice traffic would still be a good match for today's business data applications. The exponential structure, although simple and may be an appropriate choice for market purposes, its efficiency is open to debate.

With the development of various telecom services, the service providers may consider some more advanced economic issues. For example, current service providers may offer diverse service for the customers, including bandwidth access, web hosting, cable TV, digital phone, etc. The service providers may bind several services together and sell them to the customers with a considerable lower price than that if the customers purchase them separately. Moreover, the user population is not uniform: they may perceive quite different value (i.e., utility) between each other when they receive the same service. The consumers may show very different willingness to pay for the services. These issues add difficulties for the service providers to determine the tiers and the prices.

1.2 Contributions

We focus on several economic issues for tiered service networks. We first consider the problem on how to choose the service tiers and their prices optimally. We formulate the problem 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. We then discuss the strategies on market segmentation and service bundle for the service provider. Our work can be applied to different scenarios, from the market with discrete user demand requests to that with stochastic user demands, from the market with one monotonous user group to that with diverse user groups, from the market that offers only one service to that offers service bundlings, etc.

1.3 Thesis Organization

This dissertation is organized as follows. Chapter 1 is the introduction chapter. Chapter 2 introduces the related work on tiered services. Chapter 3 discusses the optimal location of service tiers for the markets with discrete user demands. Chapter 4 proposes an economic model for the markets with stochastic user demands, including the optimal location of service tiers and price bargaining model. Chapter 5 and Chapter 6 discuss the strategies of market segmentation and service bundle, respectively. Chapter 7 summaries the dissertation and gives the directions of future work.

Chapter 2

Tiered Services Network

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 x_i denote the bandwidth request (i.e., *demand*) of user i, and define the demand vector $X =< x_1, x_2, \cdots, x_n >$, which labels the user demands in non-decreasing order of requested bandwidth, $x_1 \leq x_2 \leq \cdots \leq x_n$. We assume that the demand vector is known to the service provider.

The network provider offers p bandwidth tiers (levels) of service, where typically p is a small integer ($p \ll n$). We define $Z =< z_1, z_2, \cdots, z_p >$ as the vector of service tiers offered by the network provider. Without loss of generality, we assume that the service tiers are labeled in non-decreasing order such that $z_1 < z_2 < \cdots < z_p$. For notational convenience, we also define the "null" service tier $z_0 = 0$.

With tiered service, a user i with bandwidth demand x_i will have to subscribe to service tier z_j such that $z_{j-1} < d_i \leq z_j$ so as to experience a QoS that meets or exceeds its requirements. Figure 2.1 shows a sample mapping from a vector of 14 bandwidth demands to a vector of 6 service tiers.

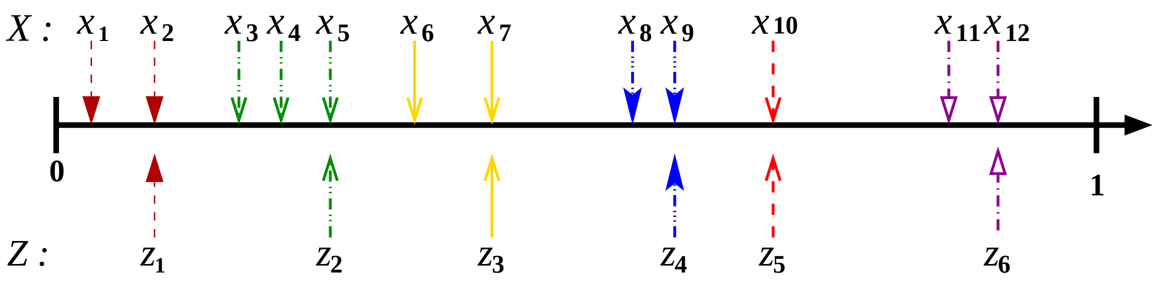


Figure 2.1: Sample mapping of bandwidth demands to service tiers

2.1 Directional p-Median Problem

Note that the network provider needs to provide each user i with additional bandwidth $(z_j - x_i) \geq 0$, and it will typically incur higher costs for doing so. Now the service provider is facing an optimization problem: How to minimize the sum of the additional cost? This is a typical directional p-median problem [15]. The directional p-median problem on the real line (i.e., DPM1) is defined as follows:

Problem 2.1.1 (DPM1) *Given a set X of n bandwidth demands, $x_1 \leq x_2 \leq \cdots \leq x_n$, find a feasible set Z of p supply points, $z_1 < z_2 < \cdots < z_p$, $1 \leq p \leq n$, which minimizes the following objective function:*

$$Obj(Z) \quad = \quad \sum_{j=1}^{p}(n_j z_j) - \rho X \quad = \quad \Phi(X,p) - \rho X \tag{2.1}$$

where n_j be the number of demand points mapped to supply point z_j, the density ρX is the amount of load requested by the original set of demand points, and the term $\Phi(X,p)$ is the load assigned to the users under the tiered service.

We can see that $Obj(Z)$ is the amount of excess bandwidth needed by the tiered service network to accomodate the demand set X after mapping it to the supply set Z. Notice that ρX is a constant, we can use $\Phi(X,p)$ to replace $Obj(Z)$.

Let $X_k = \bigcup_{i=1}^{k} x_i$, $k = 1, 2, \cdots, n$ is the set with the k smallest demand points in X. The DPM1 problem can be solved by the following dynamic programming algorithm [15].

$$\Phi(X_1, l) \quad = \quad x_1, \quad l = 1, \cdots, p \tag{2.2}$$

$$
\begin{aligned}
\Phi(X_k, 1) &= kx_k, \quad k = 1, \cdots, n && (2.3) \\
\Phi(X_k, l+1) &= \min_{q=l,\cdots,k-1} \{\Phi(X_q, l) + (k-q)x_k\} && (2.4) \\
& l = 1, \cdots, p-1;\ k = 2, \cdots, n && (2.5)
\end{aligned}
$$

The running time complexity of this dynamic programming is $O(pn^2)$. By exploiting the fact that the $n \times p$ maxtrix Φ satisfies the concave Monge condition [1, 31], an improved algorithm with time complexity $O(pn)$ is developed in [31], which is very efficient when n is large.

The directional p-Median problems in multi-dimensional space are proved to be NP-Complete [17]. Thus it is unlikely to develop a polynomial-time algorithm to find the optimal solution for such problems. The common method to solve directional p-Median problem in two dimension space (DPM2) is TB heuristic [35] and DH heuristic [16]. Both of the two methods randomly choose p points from the candidate sets as the seed, and use tremendous iterations to replace each of the p points with a better point in the candidate set. The iteration continues as a loop and it stops until the objective value converge to a constant value. The main difference between these two methods is that DH heuristic uses a smaller candidate set than the TB heuristic does. Thus, TB heuristic has lower computational cost, which is at the expenses of accuracy,

2.2 TDM Emulation

A special version of tiered service is that the service tiers are multiples of a basic bandwidth unit. The operation of the network with such a set of service tiers is similar to the time division multiplexing (TDM) networks, such as PSTN, SONET/SDH and GSM network. For example, Legacy telephone networks and SONET/SDH network have the rigid requirements that hierarchy of rates should be multiples of a voice channel (i.e., 64Kbps), resulting a proliferations of data services, such as DS-1 (1.5Mbps), DS-3 (45Mbps) and STS-3 (155Mbps) and STS-48 (2.5Gbps), etc. The term *TDM Emulation* is used to describe a tiered service network that allocates bandwidth in arbitrary multiples of a basic bandwidth unit [32, 30]. Such techniques are used in several network contexts including the next generation SONET/SDH networks and traffic grooming [9]. Fig. 2.2 shows an example of

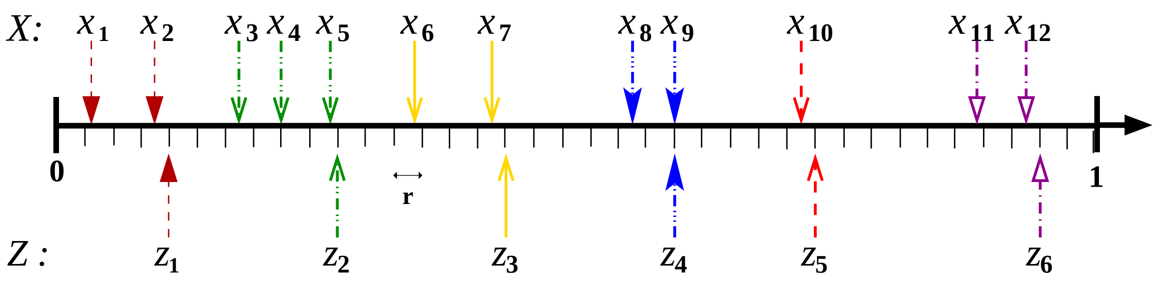


Figure 2.2: Example of service tier mapping in TDM emulation

service tier mapping in TDM emulation: 12 demand points are mapped to 6 tiers and the tiers are in units of r.

The objective in the tiered-service network with TDM Emulation is to select jointly the unit rate r and the service tiers in some optimal manner. It is expressed as follows [32]:

Problem 2.2.1 (TDM) *Given a set X of n bandwidth demands, $x_1 \leq x_2 \leq \cdots \leq x_n$, and a constant C as the maximum link bandwidth, find a real r and a feasible set Z of p supply points, $z_1 < z_2 < \cdots z_p$, $1 \leq p \leq n$, which minimizes the following objective function:*

$$Obj(Z,r) = \{\sum_{j=1}^{p}(n_j z_j) - \rho X\} + \frac{B}{r} \tag{2.6}$$

under the constraints: $z_j = rk_j$, k_j: integer, $j = 1, \cdots, p$

We can see that the TDM problem is a special case of DPM1. Consider an instance of TDM problem in which the value of r is fixed at $r = r'$, it can be solved in time $O(pK)$, where $K = \lceil \frac{x_n}{r'} \rceil$ and x_n is the largest demand.

In the objective function, the two items - one in the brackets and one not - represents a tradeoff with respect to the selection of r. Apparently, the item in brackets becomes smaller as r decreases; it is minimized when $r = \frac{1}{C}$, i.e., when bandwidth is allocated at the finest possible granularity. Increment of r will yield a suboptimal solution due to the excess bandwidth penalty expressed by the term in brackets.

The second term $\frac{B}{r}$ is introduced to prevent r from being very small. Here the constant B is related with some overhead operation of the system. To illustrate, let us make the simplifying assumption that all users request and receive the basic rate of rC

bits/sec. During the switch of service among the users, The system will do some overhead operation such as the bookkeeping operations, memory lookups, etc. Let α be the number of bits that could be transmitted during this time interval at the given service rate. The quantity $\frac{\alpha}{rC}$ represents the amount of overhead time relative to the unit rate. Let $B = \frac{c\alpha}{C}$ denote the overhead, where c is a constant which ensures that the overhead and $Obj(Z, r)$ are expressed in the same units.

In TDM problem, one important issue is how to choose the value of r. The objective of TDM is to determine the value of r so as to strike a balance between the two conflicting objectives. One way is to set a basic bandwidth unit r_0, then let $R = r_1, r_2, \cdots, r_k$, where $r_k = kr_0$, $k = 1, \cdots, K = \lceil \frac{C}{r} \rceil$. Try all the value of r_i we can have an optimal value of r.

Chapter 3

Optimal Sizing of Tiered Service Networks

In Chapter 2, the optimization goal for the service providers is to minimize the extra bandwidth introduced by mapping each demand point to the tier with a higher rate. Though the generated tiers lead to minimal extra bandwidth, they may not necessarily result in the most profit (i.e., the total revenue minus the cost). Thus, these tiers may not meet the needs of the ISPs who are always pursuing more profits. Besides, these tiers may also not satisfy the customers, who prefer to maximize the surplus of the utility minus the price. The word 'utility' is a economic term which is a measure of the value that users receive from the service; it also stands for the customers' willingness to pay for the service. Considering these economic issues, we need to develop new technologies to determine the service tiers locations.

In this chapter, we study how to optimally locate the service tiers (i.e., sizing of the network) from three perspectives: one that considers the interests of the customers only, one that considers the interest of the service provider, and one that considers both simultaneously [20].

3.1 Service Tier Optimization

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 x_i denote the bandwidth request (i.e., *demand*) of user i, and define the demand vector $D =< x_1, x_2, \cdots, x_N >$, where we have labeled the user demands in non-decreasing order of requested bandwidth, $x_1 \leq x_2 \leq \cdots \leq x_N$. We assume the demand vector is known to the service provider [1]

The network provider offers K bandwidth levels (tiers) of service, where typically K is a small integer such that $K \ll N$. We define $Z =< z_1, z_2, \cdots, z_K >$ 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 $z_1 < z_2 < \cdots < z_K$. For notational convenience, we also define the "null" service tier $z_0 = 0$.

As we mentioned in Chapter 2, with tiered service, a user i with bandwidth demand x_i will have to subscribe to service tier z_j such that $z_{j-1} < x_i \leq z_j$ so as to experience a QoS that meets or exceeds its requirements. Note that the network provider needs to provide each user i with additional bandwidth $(z_j - x_i) \geq 0$, and will typically incur higher costs for doing so; consequently, the provider will be inclined to select the service tiers, and the corresponding price to charge, so as to recoup these costs (and make a profit). On the other hand, user i subscribes to a service (i.e., z_j) that is at least as good as the one requested (i.e., x_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 trade-offs, and to develop techniques to select the service tiers and prices in a manner that accounts for both the users' and providers' perspectives.

Consider now the demand-supply relationship between the users and 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* [4, 10], 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

[1] As we will discuss in Chapter 4, our work can be extended to the case where only information about the *probability distribution* of the user demands, not the exact demand vector, is known.

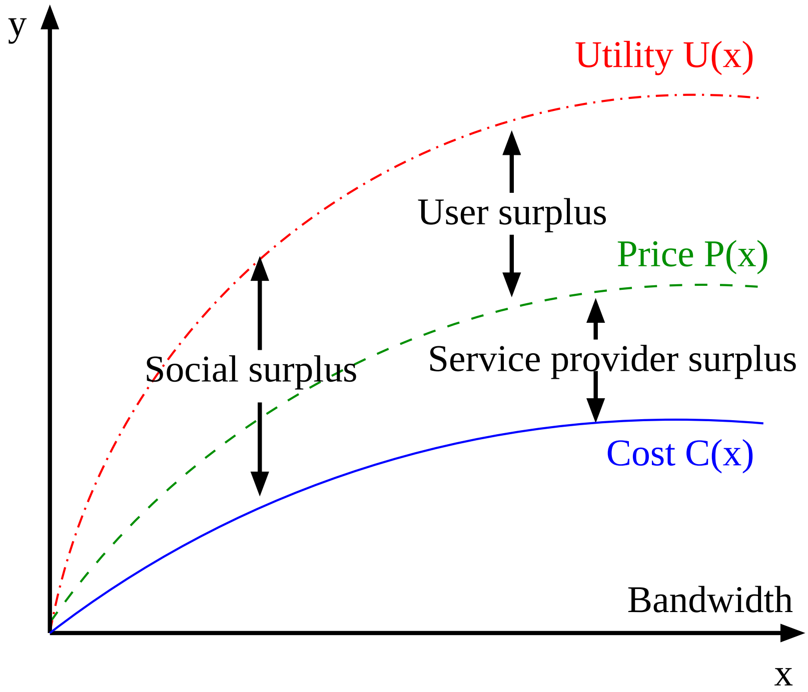


Figure 3.1: Utility, cost, and price functions

the *service provider surplus* [4, 10], defined as the difference between price and cost.

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 service provider surplus. We will refer to the social welfare as *social surplus* [4, 10]. 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, as we explain in the next chapter.

We assume the existence of three non-decreasing functions of service x: 1) the *utility* function represents the user's utility from the service; 2) the *cost* function, $C(x)$, represents the cost incurred by the provider for offering the service; 3) the *price* function, $P(x)$, represents the amount that the service provider charges for the service. Figure 3.1 shows the three function curves and the three kinds of surplus.

From Figure 3.1 we can see 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). In this dissertation 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.

Let us define the user surplus $S_{usr}(x) = U(x) - P(x)$, the provider surplus $S_{pr}(x) = P(x) - C(x)$, and the social surplus $S_{soc}(x) = U(x) - C(x)$. In the tiered-service network under consideration, the problems of maximizing the surplus of users, service providers, or society, amount to selecting appropriately the set of service tiers to be offered, as we discuss next.

3.1.1 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 in the next chapter. 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 3.1.1 (STO-U) *Given a vector X of N bandwidth demands, $x_1 \leq \cdots \leq x_N$, an integer number $K < N$ of service tiers, a utility function $U(x)$, and a price function $U(x)$, find a service tier vector $Z =< z_1, \cdots, z_K >$ that maximizes the objective function (aggregate user surplus):*

$$F_U(Z) \quad = \quad \sum_{j=1}^{K} \mid X_j \mid (U(z_j) - P(z_j)) \tag{3.1}$$

subject to the constraints:

$$z_{j-1} \quad < \quad x_i \quad \leq \quad z_j, \quad x_i \in X_j, \; j = 1, \cdots, K \tag{3.2}$$

where X_j is the set of demands mapped to service tier $z_j, j = 1, \cdots, K$, and $z_0 = 0$ is the "null" service tier we defined earlier.

The STO-U problem is an instance of the directional k-median problem DPM1 we introduced in [31], 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 (3.1) 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:

$$\Phi(n,1) = n(U(x_n) - P(x_n)), \quad n = 1, \cdots, N \tag{3.3}$$

$$\Phi(1,k) = U(x_1) - P(x_1), \quad k = 1, \cdots, K \tag{3.4}$$

$$\Phi(n, k+1) = \max_{q=k,\cdots,n-1} \{\Phi(q,k) + (n-q)(U(x_n) - P(x_n))\}$$
$$k = 1, \cdots, K-1;\ n = 2, \cdots, N \tag{3.5}$$

Expression (3.3) states that when $K = 1$, due to constraints (3.2), the optimal service tier is equal to the largest demand x_n; in this case, all n users obtain utility $U(x_n)$ and pay price $P(x_n)$. Expression (3.4) states that if there is only one user with demand x_1, the user will select a service tier equal to x_i[2]. The recursive equation (3.5) can be explained by noting that the $(k+1)$-th service tier must be equal to the largest demand x_n. If the k-th service tier is equal to x_q, $q = k, \cdots, n-1$, the aggregate user surplus is given by the expression in brackets in the right-hand side of (3.5), 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 (3.3)-(3.5) takes time $O(KN^2)$. By exploiting the fact that the $N \times K$ matrix Φ satisfies the concave Monge condition [2], 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 [31].

[2]Strictly speaking, expression (3.4) will optimize the objective function (3.1) only when the gap $U(x) - P(x)$ between the utility and price functions does not increase for $x > x_1$. 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 (3.4) in the dynamic programming solution. More generally, let $x^\star > x_1$ be a value such that $U(x^\star) - P(x^\star)$ is maximum over all $x > x_1$; then, one would set $\Phi(1, k)$ in (3.4) to $U(x^\star) - P(x^\star)$. We have not considered this alternative in this work.

3.1.2 Service Tier Optimization: The Provider Perspective

The goal of the service provider is to maximize its aggregate surplus [4]), $P(x) - C(x)$, 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 (3.3)-(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[3] 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 3.1.2 (STO-P) *Given a vector D of N bandwidth demands, $x_1 \leq \cdots \leq x_N$, a price function $P(x)$, a bandwidth cost function $C(x)$, and a per-service tier cost α, find the number $K, K \leq N$, of service tiers and an optimal service vector $Z =< z_1, z_2, \cdots, z_K)$ that maximize the objective function (aggregate provider surplus):*

$$F_P(K, Z) \quad = \quad \left\{ \sum_{j=1}^{K} \mid X_j \mid (P(z_j) - C(z_j)) \right\} \; - \; \alpha K \tag{3.6}$$

subject to the constraints:

$$z_{j-1} \;\; < \;\; x_i \;\; \leq \;\; z_j, \quad x_i \in X_j, \; j = 1, \cdots, K \tag{3.7}$$

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

---

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

The objective function (3.6) 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:

$$F'_P(K, Z) = \sum_{j=1}^{K} \mid X_j \mid P(z_j) - C\left(\sum_{j=1}^{K} \mid X_j \mid z_j\right) - \alpha K \tag{3.8}$$

The STO-P optimization problem is a generalization of the directional k-median problem [31] as the number K of service tiers is now a variable. STO-P belongs to the family of uncapacitated facility location problems [8]. 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 (3.6) that is based on dynamic programming; the STO-P problem with the alternative objective function (3.8) 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 (3.6) when the number of users is n and the number of service tiers is $k \leq n$. We can obtain $\Phi(n, k)$ using the following recursion:

$$\Phi(n, 1) = n(P(x_n) - C(x_n)) - \alpha, \quad n = 1, \cdots, N \tag{3.9}$$

$$\Phi(1, k) = P(x_1) - C(x_1) - \alpha k, \quad k = 1, \cdots, N \tag{3.10}$$

$$\Phi(n, k+1) = \max_{q=k,\cdots,n-1} \{\Phi(q, k) + (n-q)(P(x_n) - C(x_n))\} - \alpha$$
$$k = 1, \cdots, N; \; n = 2, \cdots, N \tag{3.11}$$

Expressions (3.9)-(3.11) are similar to expressions (3.3)-(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 (3.11), 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 (3.4) apply to expression (3.10) 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 $\Phi(N,k^\star) \geq \Phi(N,k)$ 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 [2], the time complexity of the dynamic programming algorithm is $O(N^2)$ using the implementation we described in [31]. 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 $O(N^2)$.

3.1.3 Service Tier Optimization: The Society Perspective

In the previous subsections we 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.

Our study is based on Welfare Economics [29, 33], a well-defined branch of Economics that aims at maximizing social welfare and fair allocation of the welfares. In tiered service networks, the social welfare (surplus) is 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 3.1.3 (STO-S) *Given a vector X of N bandwidth demands, $x_1 \leq \cdots, \leq x_N$, a utility function $U(x)$, a bandwidth cost function $C(x)$, and a per-service tier cost α, find the number $K, K \leq N$, of service tiers and an optimal service vector $Z =< z_1, z_2, \cdots, z_K)$ that maximize the objective function (social surplus):*

$$F_S(K, Z) \quad = \quad \left\{ \sum_{j=1}^{K} \mid X_j \mid (U(x_j) - C(x_j)) \right\} \; - \; \alpha K \tag{3.12}$$

subject to the constraints:

$$z_{j-1} \quad < \quad x_i \quad \leq \quad z_j, \quad z_i \in X_j, \; j = 1, \cdots, K \tag{3.13}$$

where X_j is the set of demands mapped to service tier $z_j, j = 1, \cdots, K$.

This problem is identical to the STO-P problem, except that in the objective function (3.12) $U(x)$ is used whenever $P(x)$ is used the objective function (3.6). Hence, STO-S can be solved in time $O(N^2)$ with the dynamic programming algorithm in (3.9)-(3.11).

After solving the STO-S problem, we obtain an optimal service vector $Z^\star =< z_1, z_2, \cdots, z_K >$ that maximizes the social surplus and depends only on the utility and cost functions provided by the users and network provider, respectively.

3.2 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 shown in Figure 3.2: 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 $N = 1000$, and we draw their demands $\{d_i, i = 1, \cdots, N\}$ from the corresponding distribution. We have obtained results similar to the ones presented here for other ranges and distributions of

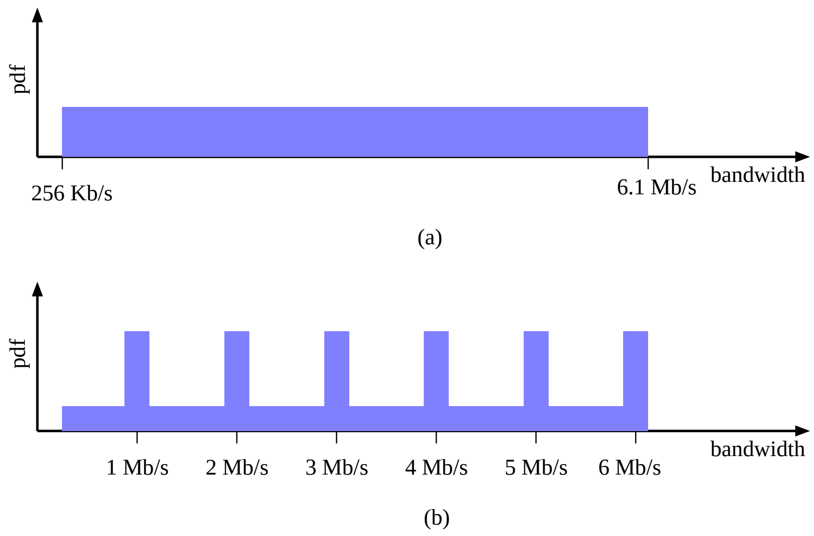


Figure 3.2: User demand distributions: (a) uniform, (b) six-modal

traffic demands and number of users; due to space constraints, we do not include these results here.

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 [34]. The parameters λ and γ can be used to control the slope of the utility function $U(x)$. In this work, we use the values $\mu = 0.5$, $\alpha = 250$, $\lambda = 10.$ and $\gamma = 0.4, 0.5$.

Due to space constraints, and since we are interested in determining the optimal prices for the service tiers, in this section we only consider the STO-S problem with the objective of maximizing the social surplus given the utility and cost functions.

3.2.1 Service Tier Selection

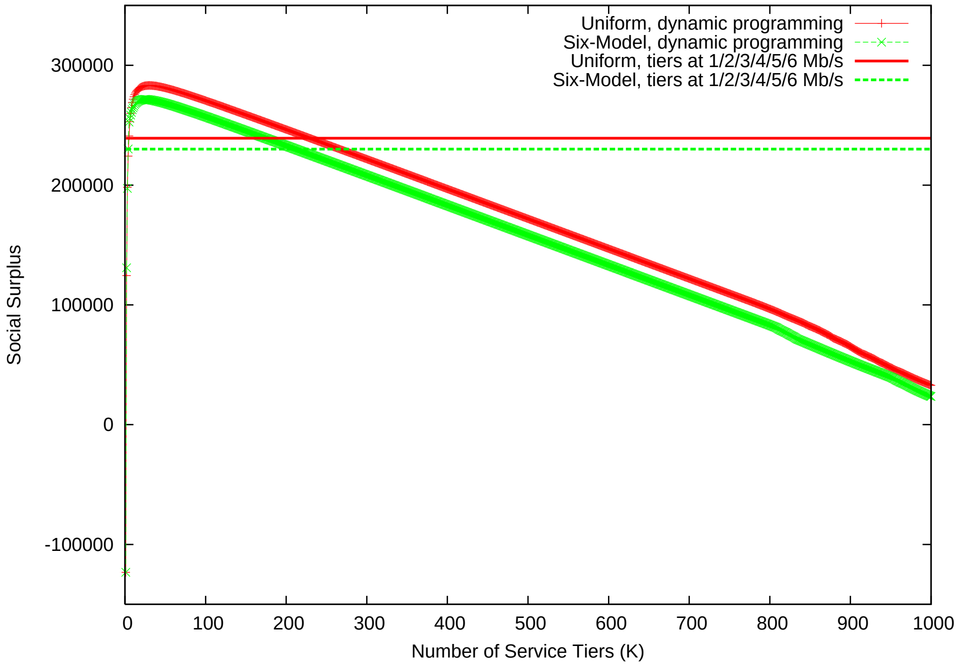
Let us first consider the impact of the number K of tiers on the value of the social surplus. Figure 3.3(a) plots the social surplus obtained by the dynamic programming algorithm against K, for the utility function with $\gamma = 0.5$. Two curves are shown, one for demands drawn from the uniform distribution and one for demands drawn from the six-modal distribution of Figure 3.2. 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; of course, the value of the optimal surplus and the number of service tiers at which the optimal is achieved does depend on the distribution of user demands, as shown in the figure.

Figure 3.3(a) 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.

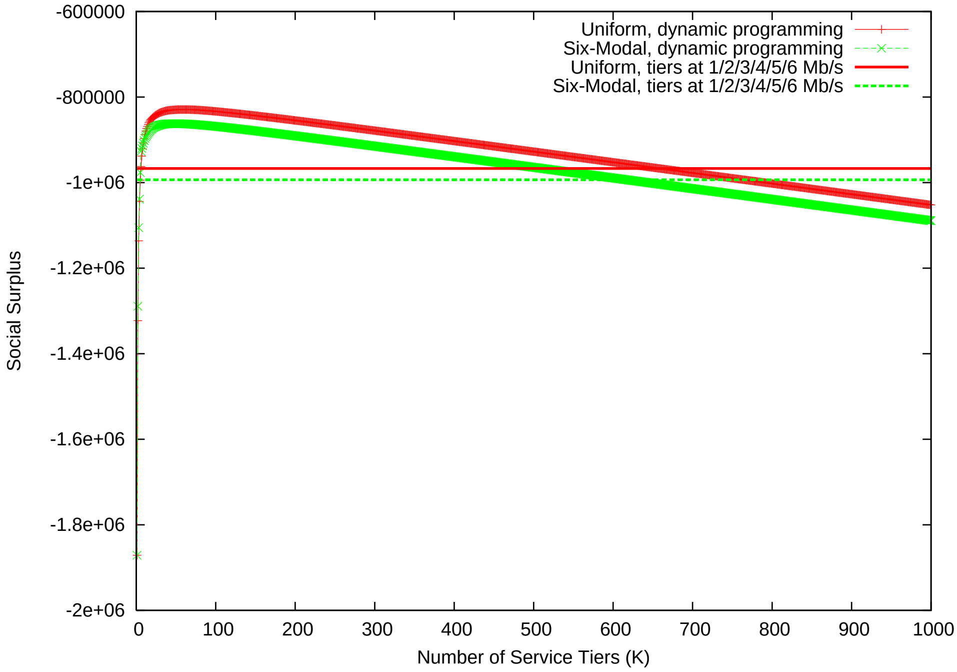
Figure 3.3(b) is similar to Figure 3.3(a), except for the fact that we used $\gamma = 0.4$ in the utility function $U(x)$. For this value of γ, the utility function is lower than the cost function $C(x)$ over most of the range of user demands. Consequently, the social surplus is negative in this case. 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.

3.3 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.



(a)



(b)

Figure 3.3: Social surplus against K: (a)$\gamma = 0.5$, (b)$\gamma = 0.4$.

Chapter 4

Economic Model for Tiered Service Networks

In Chapter 3, we use dynamic programming algorithm to search the optimal service tiers that maximizes the customer/provider/society surplus. One assumption in the chapter is that the service provider already has the bandwidth request data of each user. Such assumption may not be practicable in reality, however, because it may be difficult for some service providers to obtain such data. Nevertheless, these service providers may still be able to estimate - from history record or survey - the probability distribution of the user service requests. Therefore, it would make sense to develop a stochastic version on optimal sizing (i.e., service tier location) in tiered service networks.

In addition, Chapter 3 does not address the price setting on each tier. Usually, the price setting on each tier can be seen as the result of a negotiation (or bargaining) process between the service provider and the customers. As the market fluctuates, the negotiated price on each tier also varies. For example, one tier's price in a buyer's market would be higher than that in a seller's market. Considering this, it is necessary to model the bargaining process to set the appropriate price.

In this chapter, we first introduce the popular pricing schemes used by current multi-tiered pricing systems. We then propose an economic model to address the above-mentioned challenges. This model illustrates a stochastic version of optimal service tier location method, and the price bargaining on each tier. This model is general in scope and may be applied to a variety of contexts, independently of whether the pricing scheme is capacity-based or usage-sensitive (see section 4.1), whether charging is at the network or session/application layers, or whether the transaction is between users and provider or between providers [21, 22].

4.1 Pricing Schemes in Tiered Service Networks

Multitiered price systems are prevalent for both business and residential Internet access, and have been mainly employed in two schemes: capacity-based pricing and usage-sensitive [19].

Capacity-Based Pricing. Capacity-based schemes relate pricing to usage by setting a price based on the bandwidth or speed of the user's connection link. This is accomplished by charging for the configuration (i.e., bandwidth) of the connection, but not the actual bits sent or received. Capacity-based pricing is the prevailing pricing policy for residential broadband Internet access services. This scheme relates to our tiered service model as follows: the service is characterized by the amount of access bandwidth, each of the service tiers $z_1, \cdots, z_K$, corresponds to a certain access speed, and users are charged based on the tier to which they have subscribed. Currently, the offered tiers are either tied to the bandwidth hierarchy of the underlying network infrastructure (e.g., T1, T3, or higher for virtual private networks) or are determined in some *ad-hoc* manner (e.g., the various ADSL tiers).

Usage-Sensitive Pricing. Usage-sensitive pricing policies charge users for the actual amount of traffic they generate. In current practice, ISPs charge a customer (e.g., a video-on-demand provider) based on their traffic volume using the *95-th percentile rule* [14, 36]. Specifically, the ISP measures the user's traffic volume over 5-minute intervals during each billing period (e.g., one month), and charges the user based on the 95-th percentile value among these measured values. Typically, ISPs have a tiered pricing structure [36] in which each of the service tiers $z_1, \cdots, z_K$, corresponds to a certain traffic volume and higher tiers are mapped to higher prices. Such a structure can be mapped to our tiered service model by considering a customer with a 95-th percentile value x such that $z_{j-1} < x \leq z_j$ as having "subscribed" to tier z_j and charging the customer accordingly[1].

[1]Note that with capacity-based pricing, the tier (e.g., access speed) to which a user subscribes does not change over time (except, for instance, when a user upgrades to a higher speed), but with usage-sensitive pricing a user may be charged according to a different tier every billing period, i.e., depending on the actual

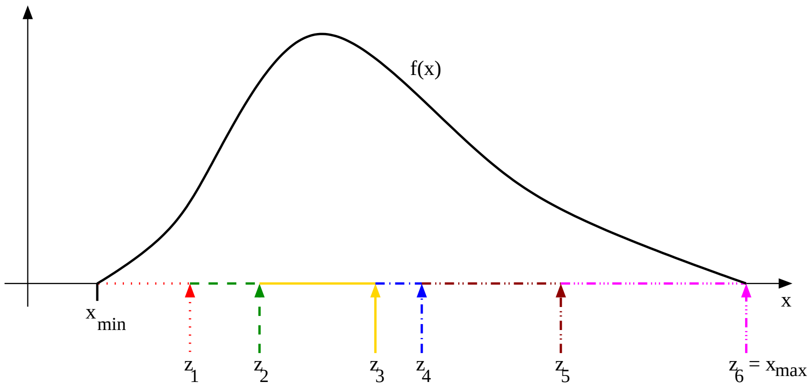


Figure 4.1: Sample mapping of service requests to service tiers

4.2 Economic Model

We consider a service characterized by a single parameter, e.g., the bandwidth of the user's access link or the amount of traffic generated by the user, and charges users on the basis of the amount of service they receive. Users may request any amount of service depending on their needs and their willingness or ability to pay the corresponding service fee. We assume that the distribution of the size x of user service requests is known; such a distribution may be obtained empirically, or extrapolated from observed user behavior and application requirements. Let $f(x)$ and $F(x)$ be the probability density function (pdf) and cumulative distribution function (cdf), respectively, representing the population of user requests. The pdf and cdf are defined in the interval $[x_{min}, x_{max}]$, where x_{min} and x_{max} correspond to the minimum and maximum, respectively, amount of service requested by any user.

The network offers K levels (tiers) of service, where typically K is a small integer, much smaller than the number of network users N (i.e., $K \ll N$). We define $Z =< z_1, z_2, \cdots, z_K >$ as the vector of service tiers offered by the network provider; without loss of generality, we assume that the service tiers are distinct and are labeled such that $z_1 < z_2 < \cdots < z_K$. For notational convenience, we also define the "null" service tier $z_0 = 0$.

traffic volume generated during each period. Nevertheless, this distinction does not affect the economic model we present in the next section.

With tiered service, a user with service request $x, x_{min} \leq x \leq x_{max}$, subscribes to service tier z_j such that $z_{j-1} < x \leq z_j$. Figure 4.1 shows a sample mapping of service requests to a vector of $K = 6$ service tiers under which all users with requests $x \in (z_{j-1}, z_j]$ subscribe to service tier $z_j, j = 1, \cdots, K$. Note also that in order to accommodate all user requests, the highest tier must be such that $z_K = x_{max}$, an assumption we will make throughout this chapter.

4.3 Optimal Location of Service Tiers

In this section, we use concepts from economics to describe the relationship between users and service providers, and we propose optimization problems of selecting the service tiers. We also illustrate how to solve these problems to obtain a set of (near-) optimal tiers. As in Chapter 3, we define the three function: utility function $U(x)$, pricing function $P(x)$ and cost function $C(x)$. We also define the user surplus $S_{usr}(x) = U(x) - P(x)$, the provider surplus $S_{pr}(x) = P(x) - C(x)$, and the social surplus $S_{soc}(x) = U(x) - C(x)$.

4.3.1 Maximization of Expected Surplus

Let $S(x)$ be a general surplus function (i.e., one of $S_{usr}(x), S_{pr}(x)$, or $S_{soc}(x)$, defined above), and suppose for the moment that the vector $Z =< z_1, \cdots, z_K = x_{max} >$ of K service tiers is given. In this case (refer also to Figure 4.1), all users with requests in the interval[2] $(z_{j-1}, z_j]$ subscribe to tier z_j, incurring a surplus of $S(z_j), j = 1, \cdots, K$. Recalling that $f(x)$ and $F(x)$ are the pdf and cdf, respectively, of user requests, the *expected* surplus $\bar{S}(z_1, \cdots, z_K)$ for the given service tier vector Z can be expressed as:

$$\begin{aligned}
\bar{S}(z_1, \cdots, z_K) &= \sum_{j=1}^{K} \left(\int_{z_{j-1}}^{z_j} S(z_j) f(x) dx \right) \\
&= \sum_{j=1}^{K} \left(S(z_j) \int_{z_{j-1}}^{z_j} f(x) dx \right) \\
&= \sum_{j=1}^{K} \left(S(z_j) \left(F(z_j) - F(z_{j-1}) \right) \right). \qquad (4.1)
\end{aligned}$$

[2] Note that the leftmost interval is $(z_0, z_1]$, where $z_0 = 0$ is the "null" service tier we defined earlier. Since $F(z_0) = 0$, the summation in expression (4.1) is correctly defined for all service tier intervals.

Consider now the problem of optimally selecting the service tiers from the users' point of view. 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 tiers so as to maximize the expected aggregate user surplus, i.e., the weighted sum of the individual user surpluses in expression (4.1) with S_{usr} in place of $S(x)$. Similarly, the goal of the service provider is to maximize its expected aggregate surplus, while considering the welfare of the society (i.e., both users and providers), the objective would be to maximize the expected aggregate social surplus. These last two objectives are obtained by using $S_{pr}(x)$ and $S_{soc}(x)$, respectively, in place of $S(x)$ in (4.1).

These three optimization problems can be formally expressed as instances of the following problem which we will refer to as the *Maximization of Expected Surplus (MAX-ES)* problem. Note that the objective function (4.2) is nonlinear with respect to the variables $z_1, \cdots, z_K$, hence MAX-ES is a nonlinear programming problem.

Problem 4.3.1 (MAX-ES) *Given the cdf $F(x)$ of user requests, an integer number K of service tiers, and a surplus function $S(x)$, find a service tier vector $Z =< z_1, \cdots, z_K >$ that maximizes the objective function (expected surplus):*

$$\bar{S}(z_1, \cdots, z_K) = \sum_{j=1}^{K} \left(S(z_j) \left(F(z_j) - F(z_{j-1}) \right) \right) \tag{4.2}$$

subject to the constraints:

$$x_{min} < z_1 < z_2 < \cdots < z_K = x_{max}. \tag{4.3}$$

Let us assume for the moment that an optimal solution to MAX-ES can be obtained; we will discuss shortly how to find such a solution. Consider the optimal solution obtained by solving the MAX-ES problem from the perspective of users or providers (i.e., by using the user $S_{usr}(x)$ or provider $S_{pr}(x)$ surplus function in place of $S(x)$, respectively). Such a solution is unlikely to be of practical value, for two reasons. First, it assumes that users and service providers may select the service tiers optimally based only on their own interests. In reality, a service tier vector that is optimal 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.

Second, both the user and provider surplus functions assume the existence of a pricing function $P(x)$. In general, the price function is the result of marketplace dynamics, including negotiation between users and service providers, hence it may not be known in advance.

On the other hand, the social surplus function $S_{soc}(x)$ depends only on the cost and utility functions, which are generally known in advance. Therefore, considering the welfare of the society as a whole overcomes the above difficulties since (1) it takes into account simultaneously the interests of both users (through the utility function) and providers (through the cost function), and (2) allows us to determine the optimal service tier vector without knowledge of the pricing function. Therefore, for the remainder of this chapter we will consider the MAX-ES problem from the society's point of view only; although, for simplicity we will continue using $S(x)$ as the surplus function, the reader should keep in mind that from now on we assume that $S(x) = S_{soc}(x) = U(x) - C(x)$.

4.3.2 Solution Through Nonlinear Programming

If the nonlinear objective function (4.2) of the MAX-ES problem is concave, and since the constraints (4.3) are convex, we may use the Karush-Kuhn-Tucker (KKT) conditions [5] to find the global maximum. The following lemma derives sufficient conditions for the function (4.2) to be concave.

Lemma 4.3.1 *If $S(x)$ and $F(x)$ are continuous and twice differentiable in $[x_{min}, x_{max}]$ and the two conditions*

$$S''(x)[F(x) - F(y)] + 2S'(x)F'(x) + S(x)F''(x) < 0 \tag{4.4}$$

$$-[S'(x)F'(y)]^2 - S(x)F''(y)\left\{S''(x)[F(x) - F(y)] + 2S'(x)F'(x) + S(x)F''(x)\right\} > 0 \tag{4.5}$$

are satisfied for all $x, y \in [x_{min}, x_{max}]$ with $y < x$, then the MAX-ES objective function $\bar{S}$ is concave in the feasible area $x_{min} < z_1 < \cdots < z_{K-1} < z_K = x_{max}$.

Proof. Define $\omega(x, y) = S(x)(F(x) - F(y))$. We can then rewrite (4.2) as:

$$\bar{S}(z_1, \cdots, z_K) = \sum_{j=1}^{K} \omega(z_j, z_{j-1}). \tag{4.6}$$

Since the sum of concave functions is also a concave function, a sufficient condition for $\bar{S}$ to be concave is for ω to be concave in the feasible area $x_{min} < y < x < x_{max}$.

The Hessian of ω is the symmetric matrix

$$\mathbf{H} = \begin{pmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \end{pmatrix}, \tag{4.7}$$

where

$$\begin{aligned} h_{1,1} &= \frac{\partial^2 \omega}{\partial x^2} = S''(x)[F(x) - F(y)] + 2S'(x)F'(x) + S(x)F''(x) \\ h_{2,2} &= \frac{\partial^2 \omega}{\partial y^2} = -S(x)F''(y) \\ h_{1,2} &= h_{2,1} = \frac{\partial^2 \omega}{\partial x \partial y} = -S'(x)F'(y) \end{aligned}$$

If ω is continuous and has continuous first and second derivatives, then it is concave if its Hessian is negative definite in the feasible area $x, y \in [x_{min}, x_{max}]$ with $y < x$, or:

$$h_{1,1} < 0 \ \text{ and } \ h_{1,1}h_{2,2} - h_{1,2}^2 > 0,$$

from which the two conditions (4.4) and (4.5) follow. ■

In general, however, the objective function may not be concave. For instance, an empirically obtained cdf F(x) may not be continuous, in which case the Hessian matrix is not defined. In such cases, it may be possible to formulate and solve approximate linear programming formulations, or apply branch-and-bound techniques [5]. One drawback of such solution methods is that they have to be customized to the specific cdf and surplus functions. More importantly, such methods may need a large number of iterations, or may get trapped at a local maximum.

4.3.3 An Efficient Approximate Solution

We now present an approximate yet efficient and accurate method for solving general instances of the MAX-ES problem. Rather than developing a sub-optimal algorithm for solving MAX-ES directly, we take a different approach: we provide an approximate formulation of MAX-ES that asymptotically converges to the formulation in (4.2)-(4.3), along with an algorithm that solves this new problem optimally.

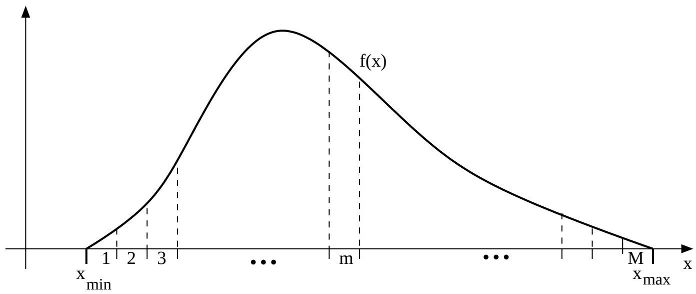


Figure 4.2: Forming a pdf approximation: The area under $f(x)$ over an interval is paired with the right-hand endpoint of the interval.

An Approximate Formulation of MAX-ES

We note that it is always possible to create a discrete approximation of the pdf $f(x)$, regardless of its form, as illustrated in Figure 4.2. In particular, we can choose an integer $M > K$ and partition the interval $[x_{min}, x_{max}]$ into M intervals each of length equal to $\frac{x_{max}-x_{min}}{M}$. The right-hand endpoint of the m-th interval is $e_m = x_{min} + \frac{m(x_{max}-x_{min})}{M}$; we associate with e_m a discrete point mass density

$$P_m = \int_{e_{m-1}}^{e_m} f(x)\, dx \, . \tag{4.8}$$

The M pairs $\{(e_m, P_m)\}$ form the approximation of $f(x)$. We also define

$$F_m = \sum_{i=1}^{m} P_i \, , \quad m = 1, \cdots, M \tag{4.9}$$

so that the M pairs $\{(e_m, F_m)\}$ form the approximation of the cdf $F(x)$.

In order to obtain an efficient solution to the MAX-ES problem, we also impose the additional restriction that the K service tiers may only take values from the set $\{e_m\}$ of the interval endpoints. Consequently, our objective is to solve the following discrete version of MAX-ES, which we will refer to as *Discrete-MAX-ES.*

Problem 4.3.2 (Discrete-MAX-ES) *Given the the M-point approximation $\{e_m, P_m\}$ of the pdf of user requests, an integer number $K < M$ of service tiers, and a surplus function $S(x)$, find a service tier vector $Z =< z_1, \cdots, z_K >$ that maximizes the objective function (approximate expected surplus):*

$$\bar{S}(z_1, \cdots, z_K) = \sum_{j=1}^{K} \left(S(z_j) \left(F_{m_j} - F_{m_{j-1}} \right) \right) \tag{4.10}$$

subject to the constraints:

$$z_j = e_{m_j} \in \{e_m\}, \quad j = 1, \cdots, K, \ m = 1, \cdots, M \tag{4.11}$$

$$z_1 < z_2 < \cdots < z_K = x_{max}. \tag{4.12}$$

Clearly, as $M \rightarrow \infty$, the pdf approximation approaches the original pdf and Discrete-MAX-ES reduces to the original MAX-ES problem.

Optimal Solution to Discrete-MAX-ES

Define $\Phi(m,k)$ as the optimal value of the objective function (4.10) when the number of intervals is m and the number of service tiers is $k \leq m$. Then, $\Phi(m,k)$ may be computed recursively as follows:

$$\Phi(m,1) = S(e_m)F_m, \quad m = 1, \cdots, M \tag{4.13}$$

$$\begin{aligned}\Phi(m,k+1) &= \max_{q=k,\cdots,m-1}\{\Phi(q,k) + S(e_m)(F_m - F_q)\}, \\ & k = 1, \cdots, K-1; \ m = 2, \cdots, M\end{aligned} \tag{4.14}$$

This algorithm is similar to that in Chapter 3. Expression (4.13) can be explained by observing that if there is only one tier of service, it must coincide with the right-hand endpoint of the m-th (i.e., rightmost) interval. The recursive expression (4.14) simply states that, for $k+1$ service tiers, the largest tier must coincide with the right-hand endpoint of the m-th interval, and the remaining k tiers must be optimally assigned to the endpoints of any feasible interval $q, k \leq q \leq m-1$.

The running time of the above dynamic programming algorithm to obtain $\Phi(M,K)$ is $O(KM^2)$. And the time complexity can be reduced to $O(KM)$ by exploiting the fact that the $M \times K$ matrix Φ satisfies the concave Monge condition [2]. Note that $\Phi(M,K)$ is the optimal solution to Discrete-MAX-ES, which in turn is an approximate formulation of the original MAX-ES problem. Hence, $\Phi(M,K)$ represents a solution close to the optimal solution to MAX-ES, regardless of the shape of the objective function (4.2). Clearly, the better the pdf approximation, i.e., the larger the value of M, the closer that $\Phi(M,K)$ will be to the true optimal solution for a given pdf; the tradeoff is an increase in running time. We have found that the value of $\Phi(M,K)$ converges quickly as the value of M approaches

50-100 for all the distribution functions we have considered, thus a (near-) optimal solution can be computed efficiently for any instance of MAX-ES.

After solving the MAX-ES problem, we obtain a service vector $Z^\star =< z_1, z_2, \cdots, z_K >$ that maximizes the social surplus and depends only on the utility and cost functions provided by the users and network provider, respectively. Next, we describe an approach to obtaining the optimal price for each service tier in $Z^\star$ in a manner that strikes a balance between the conflicting objectives of users and providers.

4.4 Optimal Pricing Based on Nash Bargaining

In this section, we apply Nash bargaining theory [26, 27] to determine optimal pricing strategies for this fixed (near-optimal) set of tiers.

Consider an optimal service vector $Z^\star =< z_1, z_2, \ldots, z_K)$ that maximizes the social surplus. We are interested in finding an appropriate price $P(z_j)$ for each service tier $z_j, j = 1, \cdots, K$, so as to satisfy both the users and service provider. Clearly, the price for each service tier z_j should be between the values of the cost and utility functions at bandwidth level z_j, as illustrated in Figure 4.3. Note that the figure plots only the component $C(x)$ of the provider's cost that corresponds to bandwidth; this is because for the given optimal service vector $Z^\star$, the number of service tiers is fixed at $K = k^\star$, where $k^\star$ is also provided as part of the solution to the MAX-ES problem. Consequently, the service tier cost, αK, is constant. In the remainder of this section, we will ignore the service tier cost and only consider the function $C(x)$; all our results can be easily adapted to account for the service tier cost by replacing $C(x)$ with function $C'(x) = C(x) + \alpha K$.

In a free telecommunication market, the price for the service is typically the result of a negotiation process between the users and service provider. This negotiation process can be modeled as a two-player bargaining game in which each player (the users or the service provider) attempts to maximize its own surplus. The outcome of the game - i.e., Nash bargaining solution - is an optimal price for the service that is mutually acceptable by both parties.

In the remainder of this section, we first give a simple introduction on Nash Bargaining, then we discuss the price bargaining on one tier and multiple tiers, respectively.

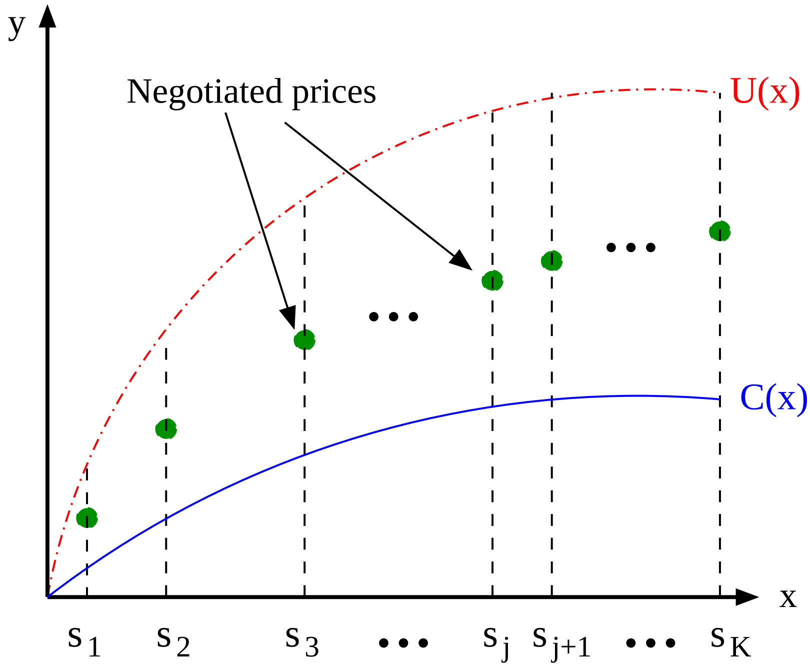


Figure 4.3: Optimal price vector

4.4.1 Nash Bargaining

Consider a bargaining between two players. A Nash bargaining game can be defined as a pair (S, d) [18], whereby $S \subset \mathbb{R}^2$ (the two-dimensional Euclidean space), is a set of the utility pairs (u_1, u_2) that the players can receive under cooperation if they reach a unanimous agreement (hence S is also called the feasible et), and $d = (d_1, d_2) \in \mathbb{R}^2$ is referred to the *disagreement point* or *threat point*, that is, the players would fail to reach an agreement if $\exists\ i \in 1, 2$ such that his utility is lower than d_i.

A Nash bargaining game assumes the following conditions: 1) S is compact and convex; 2) $d \in S$; and 3) for some $u \in S$, $u_j > d_j$ $(j \in 1, 2)$. Let B be the set of all bargaining games satisfying these three conditions. The Nash Bargaining Solution (NBS) is a function $f : B \rightarrow \mathbb{R}^2$ such that for every $(d, S) \in B$, $f(d, S) \in S$.

Nash [26] specifies four axioms that a fair bargaining solution should satisfy:

- *A1. Invariant to equivalent utility representations*: If we transform a bargaining problem (S, d) into (S', d') by taking $s'_j = \alpha_j s_j + \beta_j$ and $d'_j = \alpha_j d_j + \beta_j$, where $\alpha_j > 0$, then $f_j(S', d') = \alpha_j f_j(S, d) + \beta_j$.

- *A2. Pareto optimality*: Any player in the game is not able to increase his utility without decreasing the other player's utility. That is to say, if $s, s' \in S$ and $s'_i > s_i$, $i = 1, 2$, then $f(S, d) \neq s$

- *A3. Independence of irrelevant alternatives*: If (S, d) and (S') are bargaining problem with $S \subset S'$, and $f(S', d) \in S$, then $f(S, d) = f(S', d)$.

- *A4. Symmetry*: If S is symmetric about the axis $u_1 = u_2$ and d is on the axis, then the solution $f(S, d)$ is also on the axis.

Nash [26] proved that there exists a *unique* solution satisfying these axioms, if and only if it takes the simple form:

$$f(S, d) = arg\ max\ (s_1 - d_1)(s_2 - d_2) \tag{4.15}$$

Here f is the Nash bargaining solution. The objective of form 4.15 is to maximize the product of the utility gains of the players, which is also called *Nash product*. Notice that the Nash product and its square root attain their maximum at the same point, the objective of the form 4.15 can also be seen as maximizing the geometric average of the utility gains of the bargainers. For the proof that the Nash bargaining solution is the only solution that satisfies the above four axioms, the readers can refer to [26, 18, 38].

Consider the case that the two players are not equal in the bargaining, i.e., axiom A4 (symmetry) is not satisfied. We can introduce a parameter $\theta \in (0, 1)$ as the bargaining power. The non-symmetric bargaining solution f_θ - which satisfies the other three axioms - is givin in [28]:

$$f_\theta(S, d) = arg\ max(s_1 - d_1)^\theta (s_2 - d_2)^{1-\theta} \tag{4.16}$$

whereby $s_j \geq d_j$, $j = 1, 2$.

Apparently, if $\theta = 0.5$, equation 4.16 reduces to equation 4.15. In the extreme cases, if $\theta \to 1$, the solution maximizes s_1 subject to $s_2 \geq d_2$, and if $\theta \to 0$, the solution maximizes s_2 subject to $s_1 \geq d_1$.

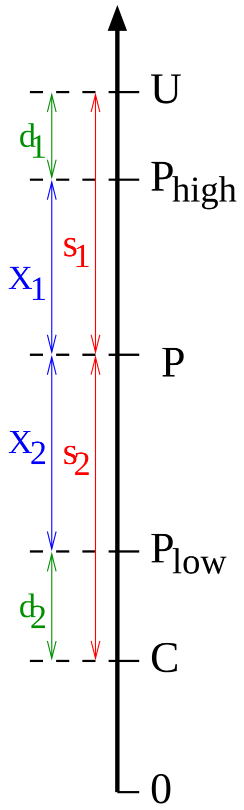


Figure 4.4: Relationship among the game parameters

4.4.2 The Single Tier Case

Let us go back to our price bargaining problem. The service provider and the customers need to determine the price in the K tiers. For simplicity, let us first consider the case $K = 1$, whereby the provider offers a single service tier z_1. For notational convenience, in this section we will simply use U, P, and C, instead of $U(z_1)$, $P(z_1)$, and $C(z_1)$, respectively.

Let P_{high}, $P_{high} \leq U$ be the maximum price that the users would accept as a satisfactory outcome of the negotiating process (game). Similarly, let P_{low}, $P_{low} \geq C$, be the minimum price that the service provider would find acceptable. Let P, $P_{low} \leq P \leq P_{high}$ be the negotiated price. We also define $X_1 = P_{high} - P$, and $X_2 = P - P_{low}$. Figure 4.4 illustrates the relationship among the game parameters U, C, P_{high} and P_{low}.

The two parties(players), users and service provider, are interested in dividing the *net* social welfare (surplus), which from Figure 4.4 is equal to $(U - C)$. The objective is to find an optimal division of the net social surplus such that both parties feel satisfied. Should the price determined, the users get the surplus $s_1 = U - P$ and the service provider gets the surplus $s_2 = P - C$. Besides, the threat point $d = (d_1, d_2)$ for the users and the service provider is $U - P_{high}, P_{low} - C$.

Let $\beta, 0 \leq \beta \leq 1$, be the bargaining power of the users, and $1-\beta$ be the bargaining power of the service provider. According to equation 4.16, we have

$$f_\theta(S, d) = arg\ max(s_1 - d_1)^\theta (s_2 - d_2)^{1-\theta} \tag{4.17}$$

$$= arg\ max(X_1)^\theta (X_2)^{1-\theta} \tag{4.18}$$

subject to the constraints:

$$X_1 + X_2 \leq P_{high} - P_{low} \tag{4.19}$$

Let $\Omega = (X_1)^\theta (X_2)^{1-\theta}$. Figure 4.5 plots the curve of the objective function Ω. The feasible area X is represented by the shaded triangle. Apparently, it is convex and bounded. The border line $X_1 + X_2 = P_{high} - P_{low}$ is the pareto optimal frontier.

As the value of Ω increases, the curve moves upwards, and vice versa. The maximum value of Ω occurs when the curve intersects the line $X_1 + X_2 = P_{high} - P_{low}$ at exactly one point, and the coordinates of this point correspond to the optimal values for X_1 and X_2.

Adding the Lagrange multiplier λ, the optimization problem is equal to

$$\Omega' = X_1^\beta X_2^{1-\beta} + \lambda(X_1 + X_2 - P_{high} + P_{low}) \tag{4.20}$$

By taking the partial derivatives of Ω' with respect to X_1 and X_2, setting them equal to zero, and using the fact that $X_1^\star + X_2^\star = P_{high} - P_{low}$, we obtain:

$$X_1^\star = \beta(P_{high} - P_{low}) \tag{4.21}$$

$$X_2^\star = (1 - \beta)(P_{high} - P_{low}) \tag{4.22}$$

From the definition of X_1 and X_2, we finally obtain the optimal prices as follows:

$$P^\star = (1 - \beta)P_{high} + \beta P_{low} \tag{4.23}$$

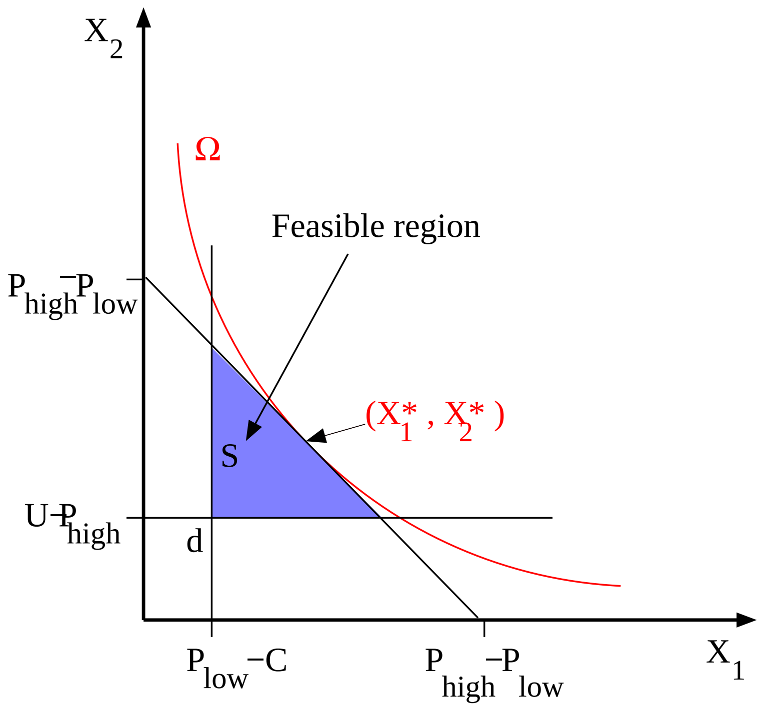


Figure 4.5: Optimal price point $(X_1^\star, X_2^\star)$

As we can see, the value of β determines the position of the optimal price along the line segment between the utility and cost values.

Let us now consider three special cases with respect to the value of the bargaining parameter β that provide some insight into the optimal solution of the above optimization problem.

Case 1 (Monopoly): $\beta = 0$, i.e., the service provider has all the bargaining power; this situation arises whenever the telecommunications market is monopolized by one service provider. In this case we have that $P^\star = P_{high}$, hence the service provider enjoys the total social surplus by squeezing out the users' surplus.

Case 2: $\beta = 0.5$, i.e., users and service provider have exactly the same bargaining power. In this case, $P^\star = 0.5(P_{high} + P_{low})$, implying that the social welfare is equally shared by the two parties.

Case 3 (Monopsony): $\beta = 1$, i.e., the bargaining power resides solely with the users; such a scenario may arise in the telecommunications market when the supply greatly exceeds the aggregate user demand. In this case we have $P^\star = P_{low}$, and the provider has to abandon any benefits (provider surplus) from supplying the service.

4.4.3 The Multiple Tier Case

Let us now consider the general case of $K > 1$ tiers of service. We can apply the methodology of the previous subsection to each service tier $z_j, j = 1, \cdots, K$, to obtain the optimal service vector $P^\star =< P^\star(z_1), P^\star(z_2), \ldots, P^\star(z_K) >$.

Using expression (4.23) we obtain:

$$P^\star(z_j) \quad = \quad (1-\beta)P_{high}(z_j) + \beta P_{low}(z_j), \quad j = 1, \cdots, K \tag{4.24}$$

Normally, the users and service providers have non-decreasing expection on their threat price (P_{high} and P_{low}) as bandwidth increases. Thus, we have that

$$P^\star(z_j) \ < \ P^\star(z_k), \quad 1 \leq j < k \leq K \tag{4.25}$$

In other words, the optimal price increases with the service tiers, i.e., with the amount of bandwidth offered to the users, consistent with intuition.

4.5 Numerical Results

To illustrate our methodology for pricing of tiered services, we consider the market for broadband Internet access under either a capacity-based or a usage-sensitive tiered pricing scheme[3].

Capacity-based pricing. We have used data collected at the San Diego Network Access Point (SDNAP) and available at the CAIDA site [11] to obtain the cdf F_{acc} of Internet access speeds shown in Figure 4.6. We adapted the raw SDNAP data so that access speeds are in the range [256 Kb/s, 12 Mb/s], typical of current broadband speeds in the United States.

Usage-sensitive pricing. We make the assumption that the monthly amount of traffic generated by users is in the range [5MB, 1TB] and follows the bounded Pareto distribution

[3] We have conducted a large number of experiments with a range of distribution, utility, and cost functions. To avoid repetition, in this study we investigate the MAX-ES problem only with the input functions described next. Nevertheless, these input functions are characteristic of real life scenarios and the results shown are representative of what we have observed in our experiments.

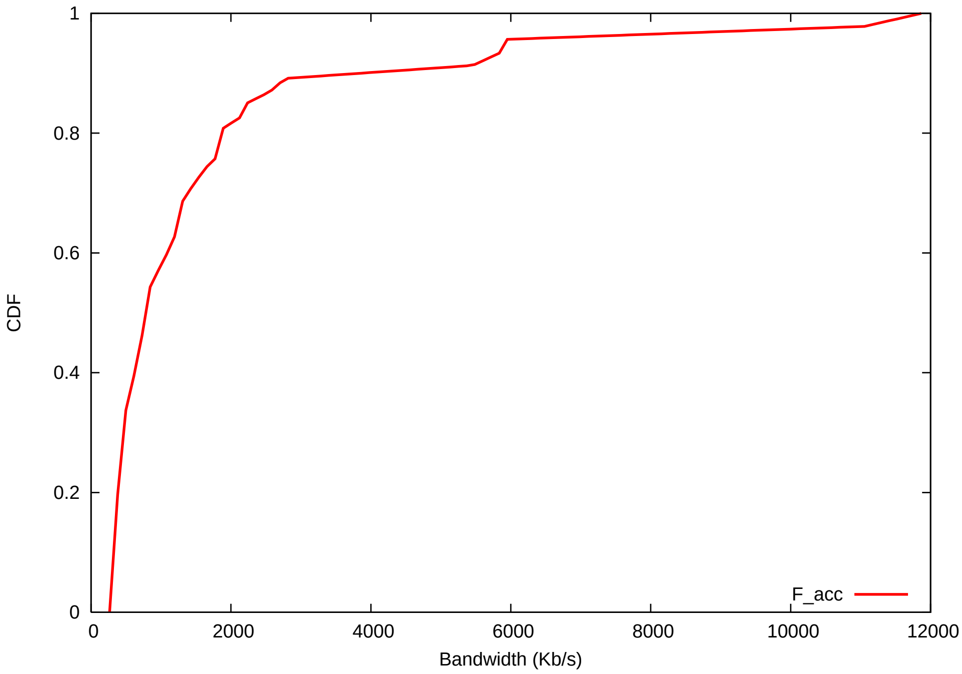


Figure 4.6: Cdf F_{acc} of Internet access speeds (data adapted from [11])

(pdf):

$$f(x) = \frac{\alpha k^{\alpha}}{1 - \left(\frac{k}{p}\right)^{\alpha}} x^{-\alpha-1}, \quad 5 = k \le x \le p = 10^6, \quad 0 < \alpha < 2 \tag{4.26}$$

We have selected two values for parameter α, corresponding to two distribution functions:

- pdf $f_{15/85}$ has $\alpha = .00001$ and is such that approximately 15% of users generate about 85% of the total traffic, and

- pdf $f_{5/50}$ with $\alpha = .03$, for which 5% of users generate approximately 50% of the overall traffic.

The latter distribution has characteristics similar to the usage patterns reported recently by one major cable ISP [12].

For all instances of the MAX-ES problem we investigate in this study, we let the utility function be

$$U(x) \;=\; \lambda x^{\gamma} \log(x) \tag{4.27}$$

and the cost function

$$C(x) \;=\; \mu x, \qquad (4.28)$$

hence the social surplus $S(x) = U(x) - C(x)$. The utility function (4.27) is an increasing, strictly concave, and continuously differentiable function of service level x, and has also been considered in the context of elastic traffic [34]. The parameters λ and γ can be used to control the slope of $U(x)$. In this work, we use the values $\lambda = 12$, $\gamma = 0.5$, and $\mu = 0.4$ for capacity-based pricing, and $\lambda = 9$, $\gamma = 0.5$ and $\mu = 0.05$ for usage-sensitive pricing, to ensure that the surplus function exhibits similar behavior across the different domains of the corresponding distributions.

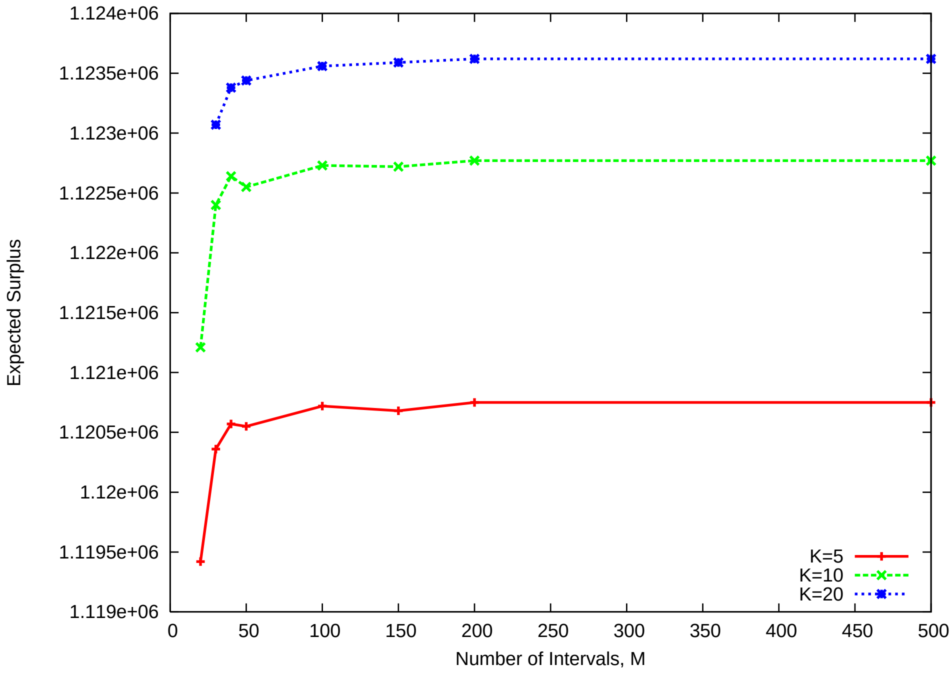
4.5.1 Service Tier Selection

Let us first consider the impact of the number M of intervals in the pdf approximation (refer to Fig. 4.2) on the convergence of the dynamic programming algorithm we presented in Section 4.3.3. Fig. 4.7(a) plots the value of the optimal solution $\Phi(M, K)$ as a function of M for the cdf F_{acc} of Figure 4.6 and the surplus function above. Fig. 4.7(b) is similar, but shows results for the Pareto cdf $F_{5/50}$.

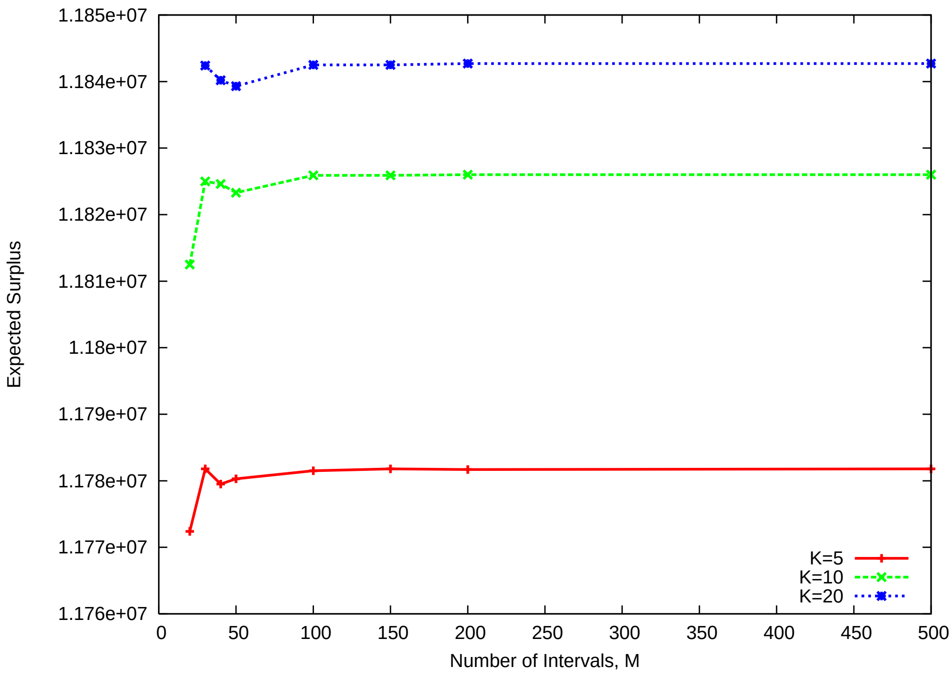
We make two important observations from these figures. First, for a given number K of service tiers, the solution obtained by the dynamic programming algorithm $\Phi(M, K)$ converges quickly as M increases. We have run experiments with a wide range of instances of MAX-ES beyond the ones we report here, and we have found that in all cases $M = 200$ is sufficient for convergence; hence we have used this value for the experiments we present in the remainder of this section. This result confirms that the dynamic programming algorithm provides an accurate and efficient solution to the MAX-ES problem.

We also observe that the expected surplus increases with the number K of service tiers. This behavior is consistent with intuition: a larger number of tiers improves the "resolution" of the final solution and allows the dynamic programming algorithm to better tailor the tiers to the given surplus and distribution functions. The figures also demonstrate the (expected) effect of diminishing returns, as further increases in the number K of service tiers provide smaller improvements to the expected surplus.

We now compare four solutions to the MAX-ES problem in terms of the expected social surplus they achieve:



(a)



(b)

Figure 4.7: Expected Surplus against M: (a)cdf F_{acc}, (b)cdf Pareto $F_{5/50}$.

1. **Optimal:** this is the optimal dynamic programming solution to the corresponding Discrete-MAX-ES instance.

2. **Optimal-rounded:** this is the tier structure derived by rounding the values of the optimal tiers of the above solution to the nearest multiple of 256 Kb/s (for capacity-based pricing) or 10 GB (for usage-sensitive pricing). The motivation for this solution arises from considerations related to marketing the service to customers who do not have intimate knowledge of the manner in which the optimal tier structure is determined. More specifically, a tier of, say, 100 GB, is likely to seem more natural and understandable to users compared to the outcome, say, 98.54 GB, of the dynamic programming algorithm, which could well be considered arbitrary.

3. **Uniform:** the K service tiers are spread uniformly across the domain $[x_{min}, x_{max}]$, i.e.,

$$z_k \quad = \quad x_{min} \; + \; k \, \frac{x_{max} - x_{min}}{K}, \quad k = 1, \cdots, K.$$

4. **Exponential:** each tier provides a level of service that is twice that of the immediately lower tier:

$$z_{k+1} = 2z_k, \quad k = 1, \cdots, K-1.$$

   As a result, the tiers divide the domain $[x_{min}, x_{max}]$ into intervals of exponentially increasing length.

The uniform and exponential are simple, straightforward solutions that do not involve any optimization and are along the lines of the structures employed by major ISPs[4]. We consider them here as baseline cases and to demonstrate that they perform poorly in terms of maximizing the expected social surplus.

Since the raw expected surplus values are not comparable across different instances of the MAX-ES problem, we introduce the concept of *normalized expected surplus* to illustrate the relative performance of the four algorithms above. For a given problem instance, let S_{max} be the maximum expected surplus value achieved by any of the four algorithms over all values of the number K of tiers evaluated in our experiments. If $\bar{S}$ is the expected

[4]Many ADSL providers offer download speeds that follow an exponential tiering structure, e.g., 384 Kb/s, 768 Kb/s, 1.5 Mb/s, 3 Mb/s, etc. Similarly for the 5/10/20/40 GB tiers of monthly traffic used in the recent pilot program by a cable ISP [12].

surplus for a given algorithm-K pair, we define the normalized expected surplus for this pair as:

$$\bar{S}_{norm} = \frac{\bar{S}}{S_{max}}, \quad 0 \leq \bar{S}_{norm} \leq 1. \tag{4.29}$$

The normalized expected surplus takes values in (0,1) for all instances of MAX-ES and provides insight into the relative behavior of the four algorithms.

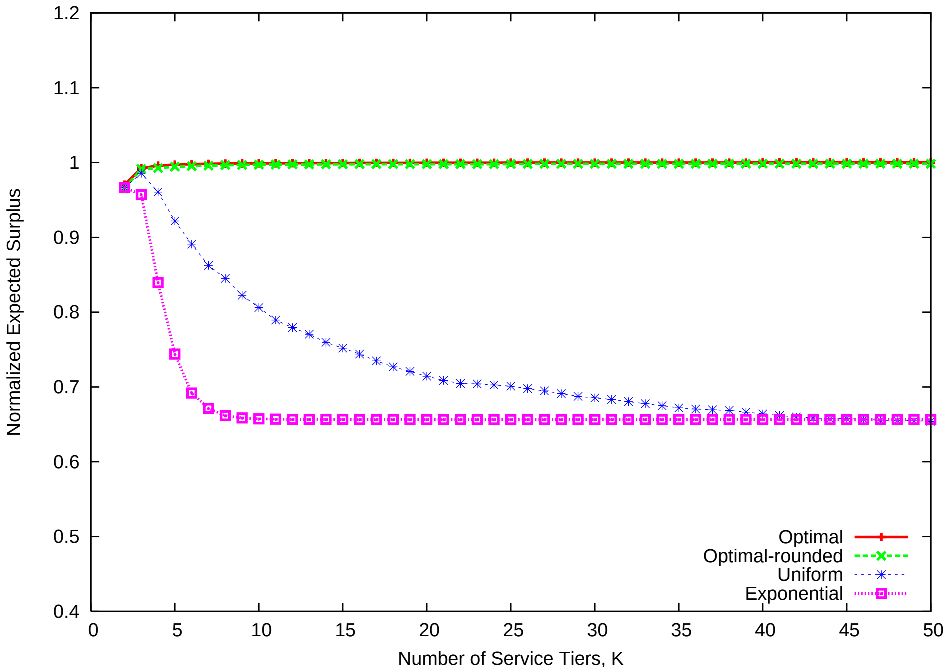


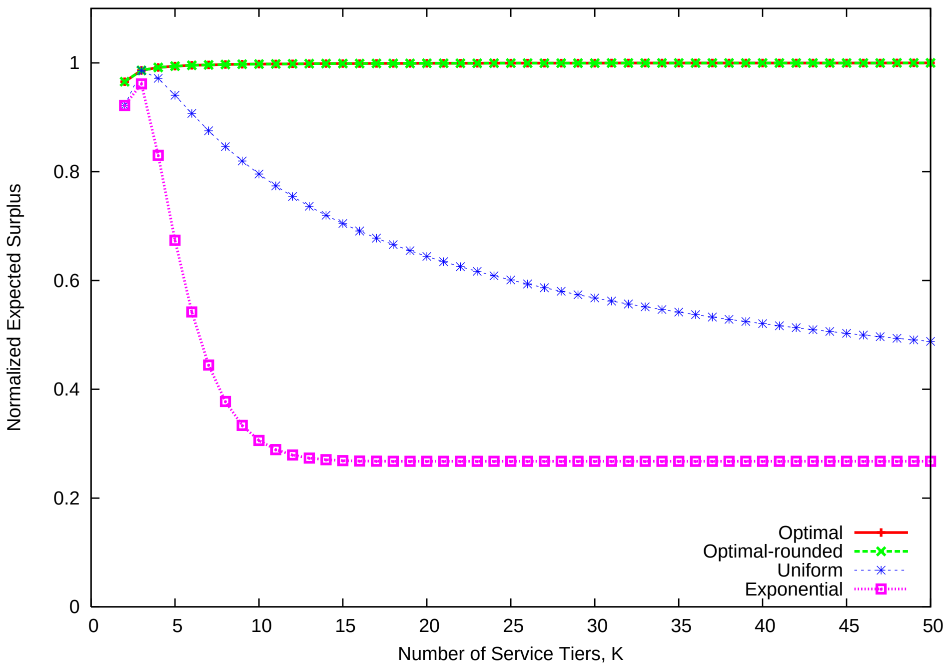
Figure 4.8: Normalized social surplus comparison, cdf F_{acc}

Figures 4.8, 4.9(a), and 4.9(b) plot the normalized expected surplus as a function of the number K of service tiers for the three distribution functions F_{acc}, $F_{5/50}$, and $F_{15/85}$, respectively. Each figure shows four curves, each corresponding to one of the algorithms for the MAX-ES problem described above. We observe that the curves for the optimal and optimal-rounded solutions almost overlap, and exhibit the best performance by far across all the values of K except very small ones, regardless of the underlying distribution function. In particular, the exponential solution decreases rapidly for $K > 2$ to about 30-50% of the optimal expected surplus, depending on the distribution (F_{acc} or Pareto). These results demonstrate that exponential grouping of customers, though favored by ISPs, performs far from optimal from an economic standpoint. In fact, the uniform tiering structure performs better than the exponential one, but it can also be far from the optimal solution for other than very small values of K.

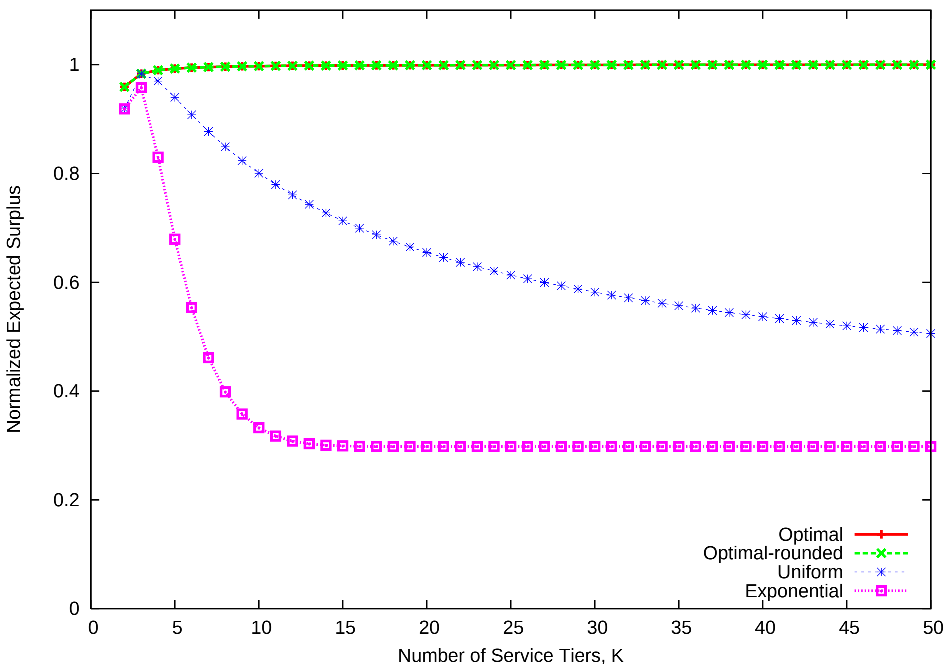
The main conclusion from the results shown in Figures 4.8-4.9(b) is that by employing our dynamic programming algorithm, which has low computational requirements, it is possible to obtain optimal tiering structures that improve the expected surplus over simple solutions by a factor of up to 2-3. More importantly, our approach makes it possible to re-optimize the tiering structures over time to accommodate evolving user demands and market conditions.

4.5.2 Optimal Pricing of Service Tiers

Figure 4.10 plots the optimal prices for the $K = 5$ optimal service tiers obtained for cdf F_{acc}. For simplicity, we let $P_{high}(z_i) = U(z_i)$ and $P_{low}(z_i) = C(z_i)$. Three price structures are shown, corresponding to the three values of the bargaining power of users $\beta = 0.25, 0.5, 0.75$.



(a)



(b)

Figure 4.9: Normalized social surplus comparison: (a)Pareto cdf $F_{5/50}$, (b)Pareto cdf $F_{15/85}$.

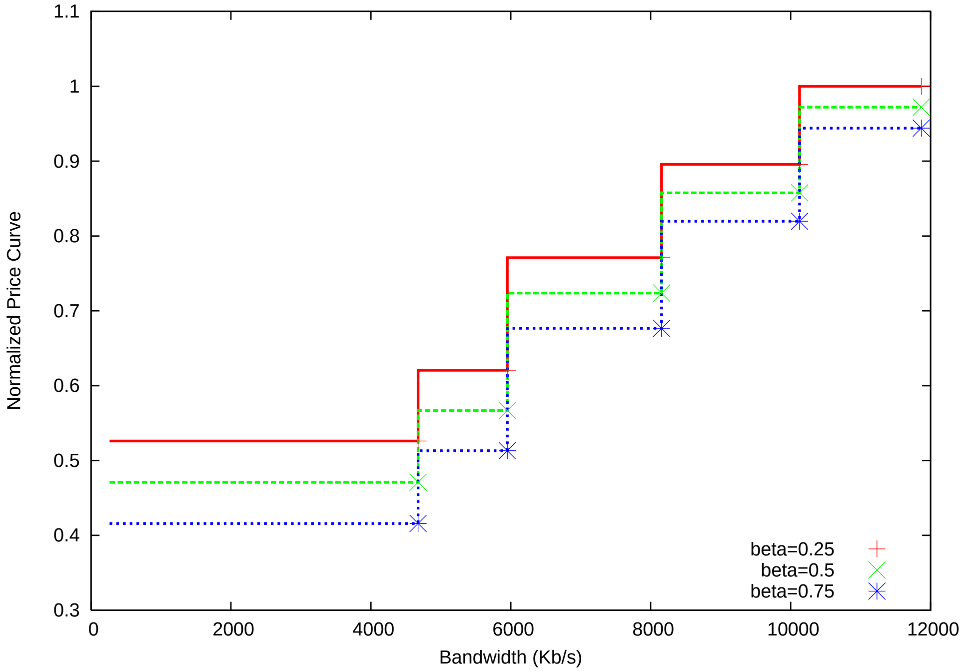


Figure 4.10: Optimal prices, cdf F_{acc}, for $K = 5$ tiers

As expected, the lower the bargaining power of users, the higher the corresponding price. Also, for a fixed value of β, the price increases with the tier index, consistent with our discussion in Section 4.4.3. Moreover, the price increase from one tier to the next is tied directly to the shape of the utility and cost functions, thus reflecting the perspective of both users and providers.

In the previous subsection we demonstrated that the exponential and uniform tiering structures are far from optimal with respect to the expected social surplus. We now show that these structures are also suboptimal in terms of the revenue collected by the service provider. Consider a tier vector $< z_1, \cdots, z_K >$, and let $P(z_j), j = 1, \cdots, K$, be the optimal price structure obtained by applying the methodology of Section 4.4. Then, the expected revenue $\bar{R}$ collected by the service provider can be calculated as:

$$\bar{R}(z_1, \cdots, z_K) \quad = \quad \sum_{j=1}^{K} \left(P(z_j) \left(F(z_j) - F(z_{j-1}) \right) \right). \tag{4.30}$$

Figures 4.11(a) and 4.11(b) plot the normalized expected revenue against the number K of service tiers for the four solutions to the MAX-ES problem we described earlier;

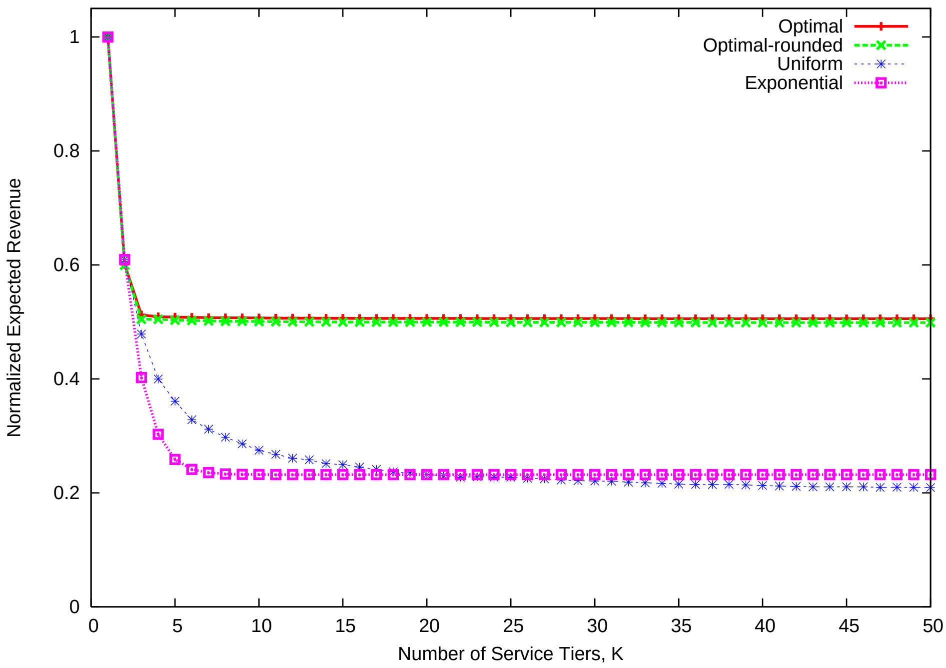
the normalized expected revenue is defined similarly to the normalized expected surplus in expression (4.29). Note that the highest revenue is obtained when there is only one tier, in which case all users are mapped to the highest possible service (that also incurs the highest price); such a solution is unlikely to be adopted in a market environment, and is included here for illustration purposes only.

As the number K of tiers increases, the expected revenue decreases for a while and then stabilizes. The curves for the optimal and optimal-rounded solutions both converge quickly to a value that is around one-half that of the maximum revenue for $K = 1$. However, the exponential and uniform solutions drop much more rapidly, eventually reaching a value that is only one-quarter (for F_{acc}) or one-sixth (for the Pareto distribution) of the maximum revenue. Again, the uniform tiering structure outperforms the exponential one, while the optimal solution achieves an expected revenue that is up to 2-3 times higher than the other two, consistent with the results of the previous section.

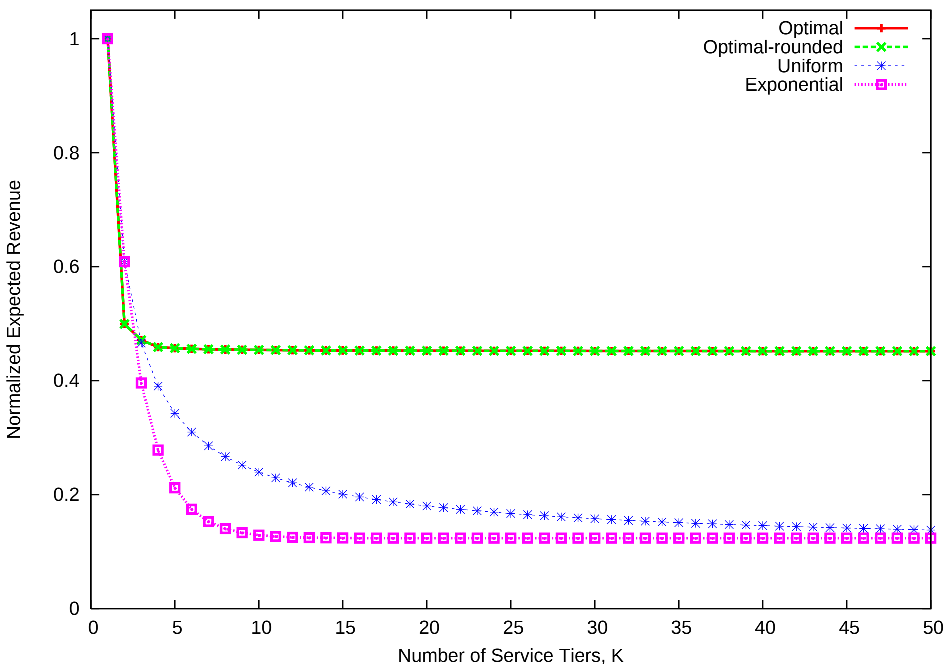
Since the cost of providing the service is the same regardless of what tiered structure is selected, these results indicate that by adopting simple, suboptimal solutions, the service provider may end up foregoing a substantial fraction of potential revenues. More importantly, these additional revenues are *not* at the expense of users, but rather due to the larger surplus achieved by the optimal solution. In other words, the optimal tiered structure provides substantial more value to both users and providers.

4.6 Extensions

The MAX-ES problem takes the number K of service tiers as input, and its objective function (4.2) is based on the assumption that there is no cost associated with offering each tier. In general, however, 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 bandwidth: the higher the access speed or the mount of traffic generated by the users, the higher the cost. We used the nondecreasing function $C(x)$ to denote this cost, which may be used to represent the link cost for carrying user traffic as well as the cost of switching the traffic in the network. The second component is due to the cost of software and hardware mechanisms (e.g., queueing structures, policing mechanisms, control plane support, etc.) required inside the network for implementing a given number K of service tiers.



(a)



(b)

Figure 4.11: Normalized revenue comparison: (a)cdf F_{acc}, (b)Pareto cdf $F_{5/50}$.

Let $C_t(k)$ be a nondecreasing function representing the cost of employing k service tiers. In this variant of the MAX-ES problem, the objective is to determine both the optimal number K of service tiers and their values so as to maximize the objective function:

$$\bar{S}(z_1, \cdots, z_K) \; - \; C_t(K) \tag{4.31}$$

where $\bar{S}(z_1, \cdots, z_K)$ is the expected surplus in expression (4.2). This problem can be solved near-optimally with the dynamic programming algorithm we presented in Section 4.3.3 after modifying the expressions (4.13)-(4.14) to account for the cost component $C_t(k)$. The running time of this algorithm is $O(M^3)$, since it has to examine all M possible values for the number of service tiers.

4.7 Concluding Remarks

In this chapter, we proposed an economic model for tiered-service networks and developed an efficient algorithm to select the service tiers in a manner that optimizes the social surplus. We also presented a method, based on Nash bargaining, to determine the optimal price for each service tier. Our approach provides insight into the selection and pricing of Internet tiered services, as well as a theoretical framework of practical importance to network providers.

Our work makes contributions in two important areas. First, we formulate the problem of selecting the service tiers (for both types of networks) 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 efficient solution approach for tackling these optimization problems. Given the set of (near-) optimal service tiers, we then employ game-theoretic techniques to find an optimal price for each service tier that strikes a balance between the conflicting objectives of users and service provider. This work provides a theoretical framework for reasoning about and pricing Internet tiered services, as well as a practical toolset for network providers to develop customized menus of service offerings. Our results also indicate that tiering solutions currently adopted by ISPs perform poorly both for the providers and the society overall.

Chapter 5

Market Segmentation

In Chapter 4, we considered a market scenario in which all users receive the same value from the service offered by the network operator, or equivalently, all users are characterized by the same utility function $U(x)$. A market in which all users value a service (or product) similarly is said to be *inelastic* [7]. Certain essential goods (e.g., gasoline or milk) and services that everyone needs tend to be inelastic, at least in the short term. Markets for most other products and services tend to be *elastic*, in that their value may be perceived quite differently across the population of consumers. Hence, in elastic markets, consumer behavior with respect to pricing may vary widely depending on the underlying utility curve that characterizes the specific consumer.

In this chapter we consider broadband Internet access as an elastic service [23]. Specifically, we assume that users are partitioned into classes, each class characterized by a distinct utility function, and we study the problem of selecting jointly the set of service tiers and their prices so as to maximize the profit (i.e., *provider surplus*).

The chapter is organized as follows. In Section 5.1 we introduce a model of user diversity. For the special case of a single tier we develop in Section 5.2 an optimal algorithm to determine both the level of service to be offered and its price. In Section 5.3 we show that introducing multiple tiers can be an effective market segmentation strategy that may lead to an increase in profits. We present performance results in Section 5.4, and we have the conclusion in Section 5.5.

5.1 Economic Model of User Diversity

We consider the market for broadband Internet access with one ISP and multiple users. The service of the ISP is described by the access speed x, with x taking values in the interval $[x_{min}, x_{max}]$, where x_{min} and x_{max} correspond to the lowest and highest speed, respectively, that the ISP may offer. The cost to the ISP of providing an amount x of service is given by the cost function $C(x)$. The ISP offers a tiered bandwidth service with p tiers. Let $Z = \{z_1, \ldots, z_p\}$ denote the set of distinct service tiers, labeled such that $z_1 < \ldots < z_p$. We also let $P(z_j)$ denote the price the ISP charges for tier z_j. Price is an increasing function of service x, hence, $i < j$ implies $P(z_i) < P(z_j)$. For notational convenience, we assume the existence of a "null" service tier z_0 for which $P(z_0) = 0$, and also that $P(z_{p+1}) = \infty$; the latter ensures that no user may receive service at an amount higher than the highest tier.

Users belong to one of T classes, $T > 1$. Users in class t are characterized by utility function $U_t(x)$. We express cost, price, and utility in the same units, e.g., US$. The various utility curves indicate the users' willingness to pay, and can be determined using market research tools such as surveys or conjoint analysis [25]. Fig. 5.1 illustrates the user diversity with respect to utility curves for $T = 4$ classes. We let $f_t, t = 1, \ldots, T$, denote the fraction of the user population that is in class t; obviously, $f_1 + \ldots + f_T = 1$. We also make the reasonable assumption that the cost $C(x)$ and utility functions $U_t(x), t = 1, \ldots, T$, are continuous, twice differentiable, and non-decreasing in the interval $[x_{min}, x_{max}]$.

If the price set for a product is *below* the value of this product to a consumer, then the consumer will purchase the product. On the other hand, if the price of the product is *higher* than the consumer's perceived value of the product, then they will not make the purchase. Consequently, users in class t will subscribe to the highest tier z_j for which the price charged does not exceed its value $U_t(z_j)$ to the users. More formally, given the utility and price functions, there is an implied mapping $h : \{1, \ldots, T\} \rightarrow Z$ from the set of user classes to the set of tiers, where $h(t) = z_j$ if and only if:

$$P(z_j) \leq U_t(z_j) < P(z_{j+1}), t = 1, \ldots, T; j = 0, 1, \ldots, p. \tag{5.1}$$

Note that, if the price of the lowest tier is higher than the utility of some class of users, then, from (5.1) these users are forced to "subscribe" to the "null" service tier z_0, which implies that they will not use the service. Fig. 5.2 illustrates the mapping of $T = 2$ classes of users to $p = 5$ service tiers based of the given price structure imposed by the step pricing function

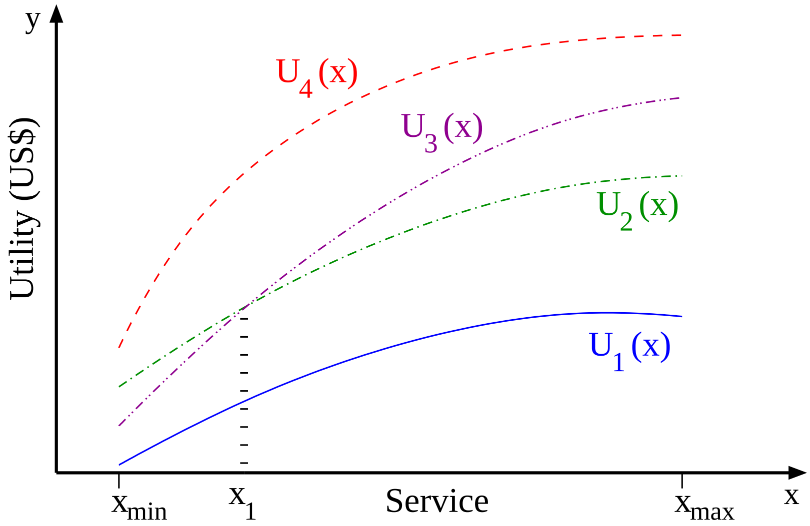


Figure 5.1: Diversity of user utility functions, $T = 4$ classes of users

$P(x)$. Specifically, users in class 1 and class 2 are mapped to tiers z_4 and z_2, respectively, consistent with expression (5.1).

From the point of view of the ISP, there is a clear tradeoff in setting the price for the service tiers. If the price for some tier is high, the ISP will lose revenue as some customers may decide to subscribe to a lower tier or not use the service at all. On the other hand, if the ISP prices the tiers conservatively, it may attract some low-utility customers, but may also forego a significant amount of revenue from customers with high utility who would be willing to pay more for the service. Therefore, the objective is to select jointly the p service tiers to be offered and their prices so as to maximize the provider surplus. We will refer to this problem as the *surplus maximization (MAX-S)* problem, and formally define it as follows.

Problem 5.1.1 (MAX-S) *Given the cost function $C(x)$, an integer T, the fraction f_t of users in class t and their utility $U_t(x), t = 1, \ldots, T$, and the domain $[x_{min}, x_{max}]$ of the cost and utility functions, find a set $Z = \{z_1, \ldots, z_p\}$ of p service tiers and their respective prices $P(z_j)$ that maximize the following objective function (provider surplus):*

$$S_{pr}(z_1, \ldots, z_p) = \sum_{j=1}^{p} [P(z_j) - C(z_j)] \sum_{h(t)=z_j} f_t \qquad (5.2)$$

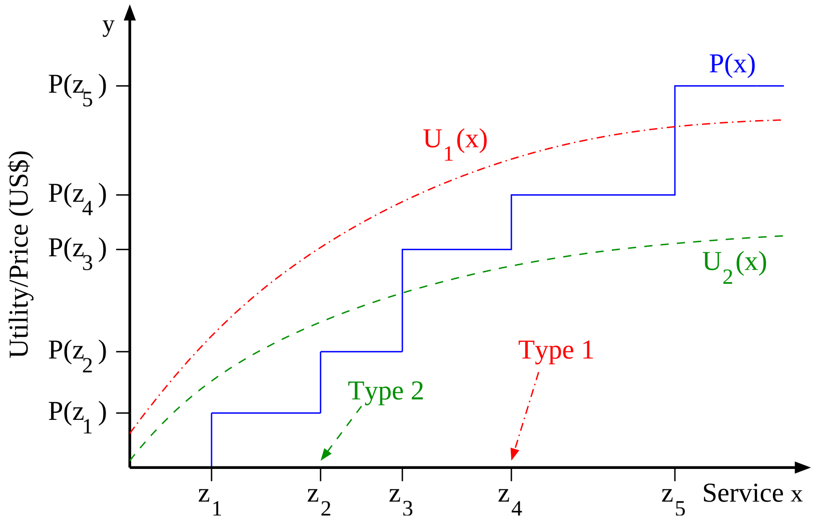


Figure 5.2: Mapping of $T = 2$ classes of users to $p = 5$ service tiers based on the price structure $P(x)$

under the constraints ($t = 1, \dots, T$; $j = 0, \dots, p$)*:*

$$h(t) = z_j \text{ iff } P(z_j) \le U_t(z_j) < P(z_{j+1}) \quad (5.3)$$

$$0 = z_0 < x_{min} \le z_1 < z_2 < \dots < z_p \le x_{max} \quad (5.4)$$

The following lemma states that the price of each service tier in the optimal solution may take one of T distinct values.

Lemma 5.1.1 *Let* $Z = \{z_1, \dots, z_p\}$ *be an optimal solution to the MAX-S problem and* $P(z_j), j = 1, \dots, p$*, be the price of the corresponding tiers. Then:*

$$\exists\, t \in \{1, \dots, T\} : P(z_j) = U_t(z_j), \quad j = 1, \dots, p. \quad (5.5)$$

Proof. By contradiction. Assume that the price of tier z_j is such that $U_s(z_j) < P(z_j) < U_t(z_j)$ for some classes $s \neq t$. In other words, class-t users (and, perhaps, users of some class q with $U_q(z_j) > U_t(z_j)$) subscribe to tier z_j, whereas users of class s and any class r with $U_r(z_j) < U_s(z_j)$ do not subscribe to tier z_j. Therefore, the price of the tier can be raised to $P'(z_j) = U_t(z_j)$ without affecting the set of users subscribing to this or any other tier. This increase in price will result in an increase to the provider surplus, contradicting the assumption about optimality of the original price $P(z_j)$. ∎

5.2 The Single Tier Case

Let us first consider the simpler case of $p = 1$, i.e., a single level of service z_1. Due to Lemma 5.1.1, we know that $P(z_1) = U_t(z_1)$ for some t, and our goal is to determine an optimal value for $z_1 \in [x_{min}, x_{max}]$ and a corresponding optimal price. To this end, we distinguish two cases.

Case 1. The T utility functions $U_t(x), t = 1, \ldots, T$, and the cost function $C(x)$ do not pairwise intersect anywhere in their domain $[x_{min}, x_{max}]$. Without loss of generality, we make the assumption that $C(x)$ lies below all of the T utility functions in the same interval. If that is not true, we can ignore the utility functions that lie below $C(x)$ and only consider the $T' < T$ functions that lie above $C(x)$. Doing so will not affect optimality, since setting the price below cost will result in a loss for the provider.

Now let us relabel the T utility functions such that:

$$C(x) < U_1(x) < U_2(x) < \ldots < U_T(x), \forall x \in [x_{min}, x_{max}],$$

and define $F_t = \sum_{s=t}^{T} f_s, t = 1, \ldots, T$, as the fraction of users falling in the classes with utilities equal to, or higher than, that of class t.

If the provider offers a single tier in the amount of z_1^t and prices it according to the corresponding utility of class-t users, then we can write the provider surplus from (5.2) as:

$$S_{pr}^t(z_1^t) \quad = \quad F_t \left[U_t(z_1^t) \; - \; C(z_1^t)\right], \quad t = 1, \ldots, T. \tag{5.6}$$

Therefore, we can find the optimal tier $z_1 \in [x_{min}, x_{max}]$ and its price using these two steps:

1. For each class t, determine the value of $z_1^t \in [x_{min}, x_{max}]$ that maximizes the quantity $S_{pr}^t(z_1^t)$ in (5.6).

2. Set the tier z_1 to the quantity z_1^q, and its price to $U_s(z_1^q)$, where q is such that $S_{pr}^q(z_1^q)$ is maximum among the T quantities computed in Step 1.

Case 2. Some of the utility and cost functions pairwise intersect in one or more points within their domain $[x_{min}, x_{max}]$. In this case, it is always possible to partition this interval into sub-intervals within which none of the functions intersect. Returning to Fig. 5.1, we can see that the domain of the utility functions can be divided into two sub-intervals, $[x_{min}, x_1]$

and $[x_1, x_{max}]$, within which the functions do not intersect. Therefore, we can obtain the optimal value for z_1 and the corresponding price by following the following steps:

1. Divide the interval $[x_{min}, x_{max}]$ into K non-overlapping sub-intervals $e_k, k = 1, \ldots, K$, such that no two utility or cost functions intersect within each sub-interval e_k.

2. For each sub-interval e_k, find the optimal value $z_1^q(k)$ and optimal price $U_q(z_1^q(k))$, as in Case 1 above.

3. Set the tier z_1 and its price to the corresponding values for the interval e_k with the maximum provider surplus $S_{pr}^q(z_1^q(k))$ among all the intervals in Step 2.

Based on the above discussion, in order to find the optimal solution to the MAX-S problem for $p = 1$ tier, we need to determine the maximum of the provider surplus function in expression (5.6) in any sub-interval $[x_1, x_2]$ of $[x_{min}, x_{max}]$. Let us define function $g(x)$ for some class t as:

$$g(x) \quad = \quad U_t(x) \ - \ C(x), \quad x \in [x_1, x_2]. \tag{5.7}$$

Since the utility and cost functions are continuous and twice differentiable throughout their domain, then function $g(x)$ is continuous and twice differentiable in any sub-interval $[x_1, x_2]$. Therefore, we can find its maximum as follows:

1. The second derivative $g''(x) \leq 0$ everywhere in $[x_1, x_2]$. Then, $g(x)$ is concave in the sub-interval, and its maximum can be found by solving the equation $g'(x) = 0$.

2. The second derivative $g''(x) \geq 0$ everywhere in $[x_1, x_2]$. Then, $g(x)$ is convex in the sub-interval, and its maximum values occur at either x_1 or x_2.

3. The second derivative changes sign in the sub-interval. In this case, we subdivide $[x_1, x_2]$ into intervals such that the second derivative g'' is either non-negative or non-positive everywhere in the smaller intervals. We obtain the maximum of $g(x)$ within each smaller interval according to either case 1 or case 2 above, from which we can select the overall maximum in $[x_1, x_2]$.

5.3 The Multiple Tier Case: Market Segmentation

Let us now return to the general case of a tiered service with $p > 1$ tiers. Such a service can be viewed as a *market segmentation* strategy [37], whereby the ISP splits the market into several segments with the goal of increasing profitability. A typical example of market segmentation is when providers offer a "premium" service at a high price for the high end of the market, and a "standard" service at a lower price for the rest of the market. An important issue that arises in the market segmentation is process is determining how to segment the market and how to differentiate among the services to be offered to the various segments so as to maximize profitability. A tiered service and price structure obtained as a solution to the MAX-S problem resolves this issue since the tiers and corresponding prices uniquely identify the market segments that optimize the provider surplus (profit).

We also note that market segmentation follows the law of diminishing returns [37] in that, after an initial increase in profits, creating an additional market segment may have a negligible effect in overall profitability. Therefore, the number p of market segments (service tiers) will, in general, be less than the number T of user classes, especially if T is relatively large. In other words, an optimal market segmentation strategy may combine several user classes into a single segment. On the other hand, because of Lemma 5.1.1, in an optimal solution to the MAX-S problem the price of each tier $z_j, j = 1, \ldots, p$, is equal to the utility $U_t(z_j)$ of some class t. Therefore, the solutions we develop are for the general case $p \leq T$.

For simplicity, in the remainder of this chapter we make the assumption that the T utility curves do not intersect anywhere in the domain $[x_{min}, x_{max}]$ and are labeled such that $U_1(x)$ is the lowest and $U_T(x)$ the highest curve. This is a reasonable assumption, since if some user A values an amount of service x_1 more than a user B, then an amount $x_2 > x_1$ of service is likely to have more value for A than for B. On the other hand, the cost function $C(x)$ may intersect with some of (or all) the utility curves within the interval $[x_{min}, x_{max}]$, but may intersect at most once with a given utility function.

5.3.1 The MAX-S Problem with Fixed Tiers

Let us first consider a restricted version of the MAX-S problem in which the p tiers are provided as input to the problem and are not subject to optimization. This problem variant arises naturally for the simple uniform and exponential tiering structures we discuss in Section 5.4. Given the minimum and maximum service levels, x_{min} and x_{max}, respectively, and the number p of tiers, the p service levels are completely specified under uniform and exponential tiering, hence only the prices of these levels need to be optimized.

Due to the assumption that utility curves do not intersect and that price is an

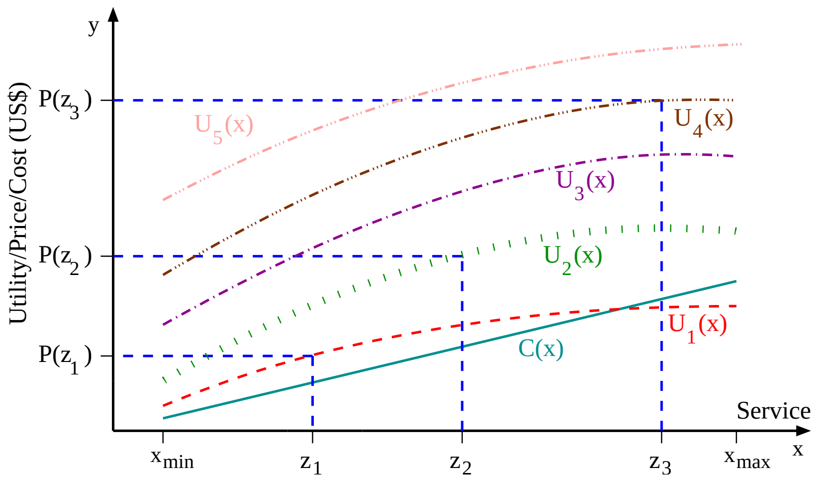


Figure 5.3: Feasible price structure for an instance of MAX-S with $T = 5$ classes of users and $p = 3$ fixed service tiers

increasing function of service x, a feasible solution to MAX-S is such that, for $j = 1, \ldots, p-1$:

$$P(z_j) = U_s(z_j) \text{ and } P(z_{j+1}) = U_t(z_{j+1}) \Rightarrow s < t. \tag{5.8}$$

as illustrated in Fig. 5.3. Therefore, we can obtain an optimal price structure for the MAX-S problem with p fixed tiers using the dynamic programming algorithm described next.

Let $\Lambda(t, l)$ denote the optimal provider surplus when there are t classes of users and l service tiers. For an instance of MAX-S with T classes and p tiers, we can compute $\Lambda(T, p)$ recursively using the following expressions (recall that the quantity F_t was defined in the previous section):

$$\Lambda(t, 1) = \max_{1 \le s \le t} \{F_s [U_s(z_1) - C(z_1)]\}, t = 1, \ldots, T \tag{5.9}$$

$$\Lambda(t, l+1) = \max_{s=l,\ldots,t-1} \{\Lambda(s, l) + F_{s+1} [U_{s+1}(z_{l+1}) - C(z_{l+1})]\}$$
$$l = 1, \ldots, p-1, \; t = 2, \ldots, T. \tag{5.10}$$

Expression (5.9) states that, if there is a single service tier fixed at z_1, the price is set to the utility that maximizes the provider surplus (refer also to the definition of provider surplus for a single tier in expression (5.6)). The recursive expression (5.10) is derived from the observation that if the price of the $(l + 1)$-th tier is set to the utility of the $(s + 1)$-th class, then all users in this class and all classes of higher utility will subscribe to this tier. The second term within the brackets in the right-hand side of (5.10) represents the contribution of this tier to the provider surplus. The first term in brackets in the right-hand side of (5.10) represents the optimal surplus for s classes of users and l tiers, $r \geq l$. Taking the maximum over all values of s yields the overall maximum. The running time complexity of this algorithm is $O(pT^2)$.

5.3.2 Approximate Solution to the MAX-S Problem

We now turn our attention to the original version of the MAX-S problem whereby both the service level at each tier and its price are subject to optimization. We solve this problem approximately by employing a discretization technique. Specifically, we divide the interval $[x_{min}, x_{max}]$ into $K > T$ segments of equal length, and impose the additional constraint that the p tiers, $z_1, \ldots, z_p$, may only take values from the set $\{e_k, k = 1, \ldots, K\}$, where $e_k = x_{min} + \frac{k(x_{max} - x_{min})}{K}$ is the right endpoint of the k-th interval. As $K \to \infty$, this discrete version of MAX-S approaches the original version in which $z_j, j = 1, \ldots, p$, are continuous variables.

Let $\Delta(k, t, l)$ denote the optimal solution to this discrete version of MAX-S when there are k points, t classes, and l tiers. We can compute $\Delta(K, T, p)$ recursively as follows:

$$\Delta(k, t, 1) = \max_{1 \leq m \leq k} \left\{ \max_{1 \leq s \leq t \leq m} \{F_s \left[U_s(e_m) - C(e_m)\right]\} \right\}$$

$$k = 1, \ldots, K, \; t = 1, \ldots, T, \; t < k \tag{5.11}$$

$$\Delta(k, t, l+1) = \max_{m = l, \ldots, k-1} \left\{ \max_{s = l, \ldots, t-1} \{\Delta(m, s, l) + \max_{r = m+1, \ldots, k} \{F_{s+1} \left[U_{s+1}(e_r) - C(e_r)\right]\} \} \right\}$$

$$l = 1, \ldots, p-1, \; t = 2, \ldots, T, \; k = 2, \ldots, K, \; t < k \tag{5.12}$$

When there is only one service tier, it is placed at some endpoint e_m and its price is set at the utility of some class s that maximizes the provider surplus, hence we have expression (5.11). In the general case of k points, t classes, and $l+1$ tiers, the optimal value is obtained by (1) considering the best placement and pricing of l tiers in $m < k$ points and $s < t$ classes, given by $\Delta(m, s, l)$, in which case the best placement and price for tier $(l+1)$ is given by the second line of (5.12), and (2) then taking the maximum over all possible values of m and s, yielding the recursive expression (5.12). The running time complexity of this algorithm is $O(pT^2K^3)$.

We have conducted a large number of experiments (omitted due to space constraints) which indicate that $K = 100$ is sufficient for the algorithm to converge; hence we use this value in the performance study we present in the next section.

5.4 Numerical Results

To evaluate the performance of tiered service as a market segmentation strategy, we consider the market for broadband Internet access. We let the minimum service x_{min} =256 Kbps and the maximum service x_{max} =12 Mbps, roughly corresponding the range of broadband speeds in the U.S. For all instances of the MAX-S problem we investigate in this study, we assume the existence of $T = 50$ classes of users characterized by the family of utility curves:

$$U_t(x) \;=\; \lambda_t x^{\gamma} \log(x), \quad t = 1, \ldots, T = 50, \tag{5.13}$$

we use the linear cost function $C(x) = \mu x$, and we set the values for parameters μ, γ and λ_t to:

$$\mu = 0.3,\ \gamma = 0.5,\ \lambda_t = 10 + .1(t-1),\ t = 1, \cdots, T = 50,$$

such that the T utility curves do not intersect in the domain [256 Kbps, 12 Mbps] and are labeled from lowest to highest.

We consider three distributions of users into classes:

- a *uniform* distribution, in which each class contains an equal fraction of the user population: $f_t = \frac{1}{T}$,
- an *increasing* distribution, such that the fraction of users in a given class increases with utility: $f_t = ct$, where $c = \frac{1}{1275}$ is a constant that ensures that $\sum_{t=1}^{T} f_t = 1$, and
- a *decreasing* distribution, in which the fraction of users in a given class increases with utility: $f_t = c(T + 1 - t)$, where $c = \frac{1}{1275}$.

5.4.1 Tier Structure Comparison

We compare the performance of four tiered structures:

1. **Optimal:** the tiered structure obtained from the dynamic programming algorithm (5.11)-(5.12).
2. **Optimal-rounded:** the tiered structure derived from rounding the values of the optimal tiers to the nearest multiple of 256 Kbps.
3. **Uniform:** the p tiers are spread uniformly across the domain [256 Kbps, 12 Mbps].
4. **Exponential:** the p tiers divide the domain [256 Kbps, 12 Mbps] into intervals that double in length (from left to right).

The uniform and exponential are simple solutions similar to structures employed by major ISPs in which the $p > 1$ service tiers are completely defined; hence, their prices were optimized using the approach we described for fixed tiers in Section 5.3.1. For the optimal structure, on the other hand, we obtained both the $p > 1$ service levels and their prices using the dynamic programming algorithm in Section 5.3.2. However, for $p = 1$, we obtained the optimal service level and its price following the algorithm in Section 5.2, and we use this value for the curves of *all four* tiering structures.

Figures 5.4, 5.5, and 5.6 plot the provider surplus against the number p of tiers and correspond to the decreasing, uniform, and increasing distribution of users into classes, respectively. Each figure show four curves, each corresponding to one of the four tiering structures. As we can see, the tiering structure (referred here as "optimal") obtained from the approximate solution to the MAX-S problem and the corresponding optimal-rounded structure outperform the uniform and exponential tiering structures across the range of values for p and across the user distributions into classes. Therefore, network providers would benefit by applying the dynamic programming solutions to determine the tiered structures to offer. Furthermore, although the uniform and exponential tiering structures uniquely define the various tiers to be offered, the prices for these tiers are determined by the dynamic programming algorithm (5.9)-(5.10) so as to optimize the provider surplus for the given tiers. Any other *ad hoc* pricing scheme would result in a lower surplus, hence even for these structures the providers would benefit from the tier pricing methodology presented earlier.

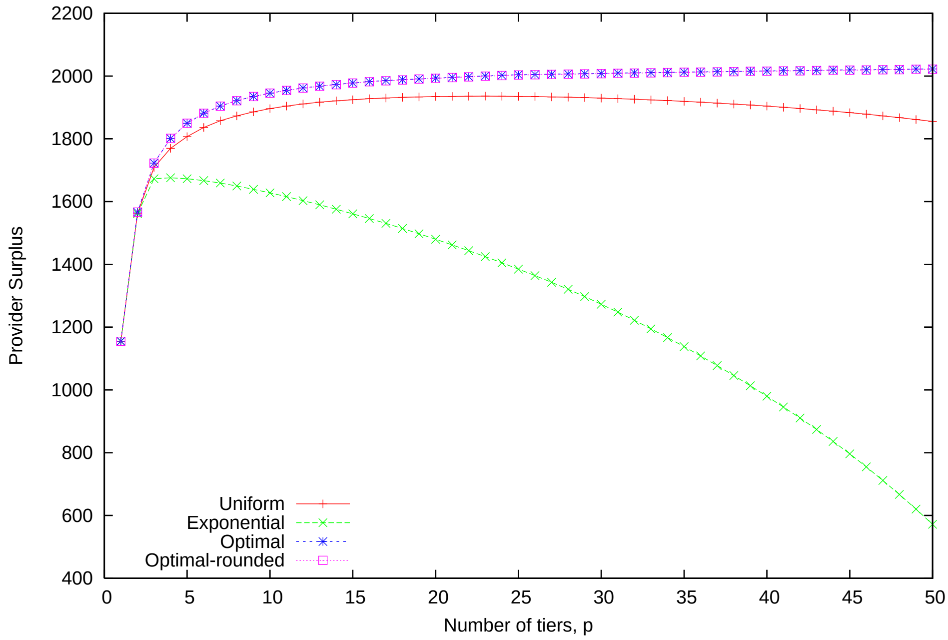


Figure 5.4: Tier structure comparison, uniform distribution of users, $T = 50$

There are several more important observations we make from these three figures. First, it is clear that the curves for the optimal and optimal-rounded structures increase rapidly with p initially, but flatten out once the number of tiers increases beyond $p = 10-15$; the latter behavior implies that additional tiers provide diminishing returns beyond a point. This result is consistent with economic theory which predicts that there is a limit to the benefits that can be achieved by segmenting a market; it also provides further confirmation to the thesis of our earlier work [31] that a relatively small number of service tiers is sufficient to capture most of the benefits of tiering.

We also observe that the exponential tiering structure performs poorly overall, and that its curves reach a well-defined maximum: the surplus achievable under such structure peaks at a small value of p and starts to decline rapidly thereafter. This behavior can be explained by noting that most of the tiers in an exponential structure are grouped together at the leftmost part of the service domain $[x_{min}, x_{max}]$ =[256 Kbps, 12 Mbps], and the few tiers that cover the remaining of the interval do not provide fine enough granularity to capture the benefits of market segmentation. The curves corresponding to the uniform tiering structure are below those for the optimal and optimal-rounded structures, but higher than those for exponential tiering. These results indicate that tiering structures with equally spaced tiers would be better for the service provider than exponential ones. Furthermore, we can see that the uniform tiering curves also reach a maximum at a certain value of p that depends on the user distribution, and start to decline as p increases further. This behavior demonstrates that simply adding more tiers but placing them into specific points in the domain of the service is not an effective market segmentation strategy; hence, to achieve the maximum benefits of market segmentation the service provider must optimize both the size and price of each tier.

Finally, we note that the overall provider surplus increases as we move from the decreasing distribution of users into classes (Fig. 5.5) to the uniform distribution (Fig. 5.4) to the increasing distribution (Fig. 5.6). This is expected, since the fraction of users characterized by high utility functions (i.e., willing to pay higher prices) is lowest for the decreasing distribution and highest for the increasing distribution. As a result, the surplus that the provider is able to extract through market segmentation is higher in the latter case.

5.5 Concluding Remarks

In this chapter, we have investigated tiered service as a market segmentation strategy for increasing ISP profits under the assumption that consumer behavior with respect to pricing varies across the user population. We developed an efficient dynamic programming algorithm for determining optimally both the service tiers and their prices. Our approach provides new insight into the selection and pricing of Internet tiered services.

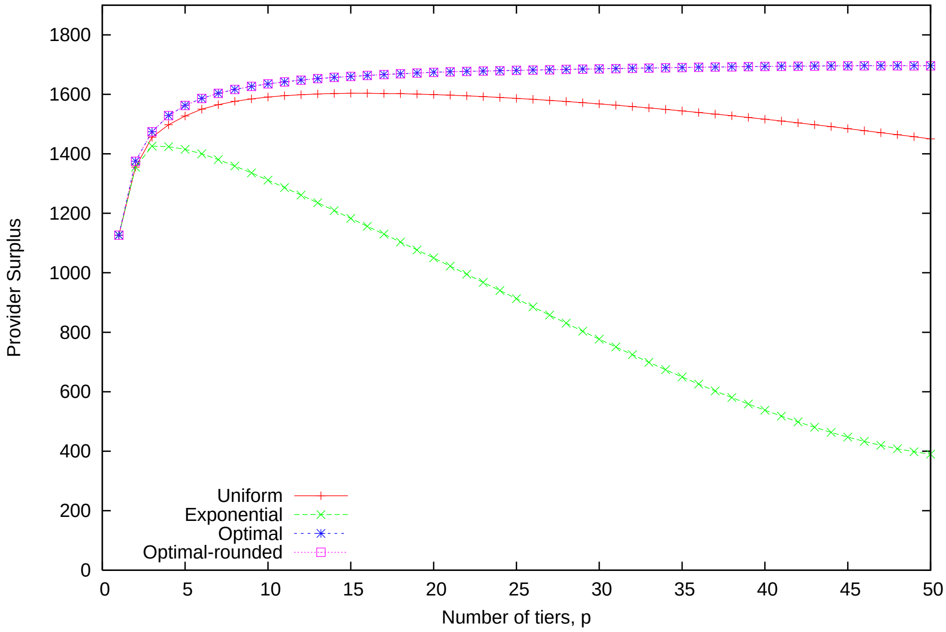


Figure 5.5: Tier structure comparison, decreasing distribution of users, $T = 50$

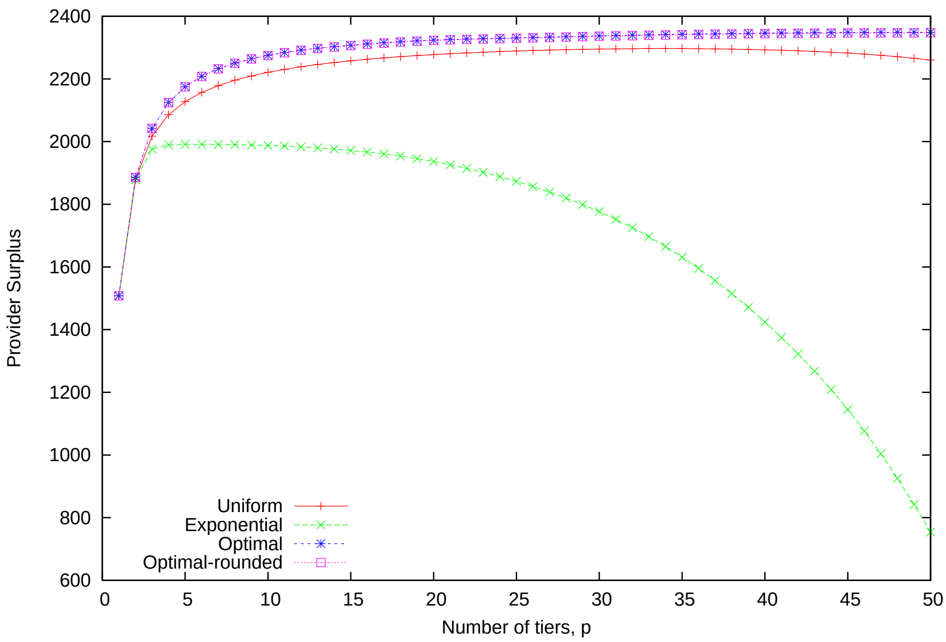


Figure 5.6: Tier structure comparison, increasing distribution of users, $T = 50$

Chapter 6

Service Bundling

In recent years, service bundling [6] has become the commonly used marketing strategy in the telecommunications market. The term *service/product bundling* means several services/products are combined together and sold as a single package. The bundles are often priced at a discount to the total price that their constituent products or services would fetch if they were sold separately.

Bundling can be beneficial to both consumers and sellers. The former, in addition to the lower overall price, may appreciate the lower transaction costs and simplified decision process compared with shopping for individual products or services, and may experience a better overall performance due to complementarities among the bundle components. For sellers, bundling has the potential to reduce production and transaction costs, reduce customer churn, and increase revenue and profitability. In particular, bundling is most successful as a marketing strategy whenever the marginal costs of bundling are low, customer acquisition costs are high, and there are economies of scale in production and distribution of the bundled products. Consequently, bundling is common in industries that share these characteristics, including the telecommunications and cable TV industry, the software business, and the fast food industry, among others.

In this chapter, we consider the problem of determining optimal tiering structures for service bundles using tools from economics and utility theory [24]. Our work provides insight into the selection and pricing of Internet tiered services. In Section 6.1 we develop an economic model for bundled network services. The optimization problem in the model is selecting jointly the tiers and their prices so as to maximize the expected profit (i.e., provider surplus [4]) of the service provider under user budget constraints. In Section 6.2 we develop dynamic programming algorithms both for the case of predetermined tiers (i.e., when only price is subject to optimization) and the general version of the problem. We present numerical results in Section 6.3, and we give the conclusion in Section 6.4.

6.1 Economic Model of Service Bundling

Consider an ISP that offers two services. One service, characterized by parameter x (e.g., access speed), may be offered at levels between a minimum x_{min} and a maximum x_{max}. The second service, say, web hosting, is also characterized by a single parameter y (e.g., corresponding to monthly amount of traffic handled), with y also taking values between a minimum y_{min} and a maximum y_{max} level. The ISP bundles the two services into a package, and offers a tiered structure with p tiers for the combined service. We let $Z = \{(z_1, t_1), \ldots, (z_p, t_p)\}$ denote the set of p distinct service tiers, where the j-th tier $(z_j, t_j), t = 1, \ldots, p$, corresponds to an amount z_j for service x and an amount t_j for service y.

We let $C(x, y)$ denote the cost to the ISP of offering a service bundle (x, y) of the two services. We also let $P(z_j, t_j), j = 1, \ldots, p$, denote the price that the ISP charges subscribers to tier (z_j, t_j). Without loss of generality, we assume that tiers are labeled such that

$$P(z_{j-1}, t_{j-1}) \quad < \quad P(z_j, t_j), \quad j = 2, \ldots, p. \tag{6.1}$$

For mathematical convenience, we also define the "null" service tier $(z_0 = 0, t_0 = 0)$ with price $P(z_0, t_0) = 0$, as well as a fictitious $(p+1)$-th service tier such that $P(z_{p+1}, t_{p+1}) = \infty$.

The value that users receive from a bundle (x, y) of the two services is described by the utility function $U(x, y)$. In essence, the utility function imposes a pairwise ranking of bundles by order of preference, where *preference* is a transitive relation. More precisely, if $U(x, y) > U(x', y')$, then bundle (x, y) is said to be strictly preferred to bundle (x', y'). On the other hand, if $U(x, y) = U(x', y')$, the two bundles are equally preferred, and the consumer is said to be *indifferent* between the two bundles. In particular, a curve

$$U(x, y) \quad = \quad u \tag{6.2}$$

is referred to as an *indifference curve* since the user has no preference for one bundle over

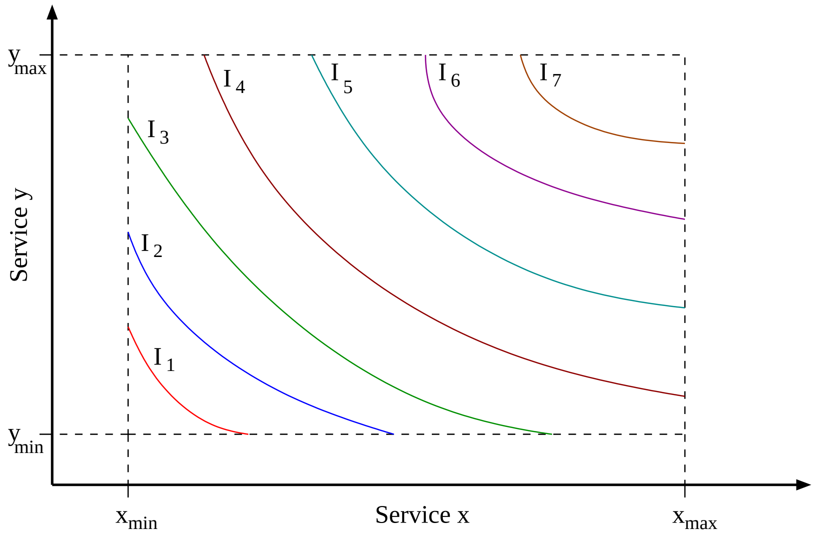


Figure 6.1: Indifference curves, $I_1, \ldots, I_7$, such that $U(x, y) =$ constant along each curve (utility is measured on the vertical z axis)

another among the bundles represented by points along this curve. In other words, each point on an indifference curve provides the same level of utility (value, or satisfaction) to the user. Indifference curves are typically used to represent demand patterns for product or service bundles observed over a population of consumers.

Fig. 6.1 shows a set of indifference curves, each associated with a different utility level. In this figure, utility is measured along the z (vertical) axis, and the indifference curves are simply the projections of the function $U(x, y) = u$, for various values of constant u, on the xy plane. In Fig. 6.1, users would rather be on curve I_7 rather than I_6; they would also rather be on curve I_6 rather than on I_5, and so on, but they do not care where they are on a given indifference curve. Indifference curves are similar to topographical maps, in that each point along a given curve is at the same "altitude" above the floor.

The characteristics of the curves in Fig. 6.1 are typical of indifference curves in general. Specifically, indifference curves are defined only on the positive quadrant of the xy plane, and they are negatively sloped and convex; in other words, as the quantity of one service or good x (respectively, y) that is consumed increases, it must be offset by a decrease in the quantity consumed of the other good y (respectively, x), so as to keep utility (satisfaction) constant.

In Economics research, the Cobb-Douglas family of functions [3] are widely used to generate indifference curves with the characteristics shown in Fig. 6.1. This parameterized family of functions is defined as:

$$U(x, y) = x^{\alpha} y^{1-\alpha}, \quad 0 \leq \alpha \leq 1, \tag{6.3}$$

where α is a parameter whose value is used to specify a certain function within the family. Then, the indifference curve for a constant level u of utility is given by:

$$y = u^{\frac{1}{1-\alpha}} x^{\frac{-\alpha}{1-\alpha}} \tag{6.4}$$

We will use the Cobb-Douglas utility in (6.3) as the utility function in this chapter.

We make the assumption that each user has a budget B, where B is a random variable defined in the interval $[B_{min}, B_{max}]$. We let $f(B)$ and $F(B)$ denote the PDF and CDF, respectively, of random variable B. We make the assumption that a consumer will make a purchase if and only if the price of the product is no greater than the consumer's budget. More specifically, given a set Z of p tiers and a price structure consistent with (6.1), a user will subscribe to the tier (z_j, t_j) with the highest index j whose price $P(z_j, t_j)$ does not exceed the user's budget B.

We are interested in selecting a set of service tiers for the bundled services, and determining their prices, so as to maximize the expected provider surplus (i.e., profit). We call this the *maximization of expected provider surplus in two dimensions (MAX-ES-2D)* problem, defined formally as:

Problem 6.1.1 (MAX-ES-2D) *Given the cost and utility functions $C(x, y)$ and $U(x, y)$, respectively, defined in the domain $[x_{min}, x_{max}] \times [y_{min}, y_{max}]$, and the CDF $F(B)$ of user budgets, find a set $Z = \{(z_1, t_1), \ldots, (z_p, t_p)\}$ of p service tiers and their respective prices $P(z_j, t_j)$ that maximizes the following objective function representing the expected provider surplus:*

$$\bar{Q}_{pr}(Z) = \sum_{j=1}^{p} ((P(z_j, t_j) - C(z_j, t_j)) \times (F(P(z_{j+1}, t_{j+1})) - F(P(z_j, t_j)))) \tag{6.5}$$

under the constraints:

$$P(z_1, t_1) < P(z_2, t_2) < \ldots < P(z_p, t_p) \tag{6.6}$$

$$P(z_j, t_j) \quad \leq \quad U(z_j, t_j), \quad j = 1, \ldots, p \tag{6.7}$$

$$x_{min} \leq z_j \leq x_{max}, \ y_{min} \leq t_j \leq y_{max}, \ j = 1, \ldots, p \tag{6.8}$$

Note that the terms $F(P(z_{j+1}, t_{j+1})) - F(P(z_j, t_j)), j = 1, \ldots, p$, in (6.5) represent the fraction of users whose budgets fall in the intervals $[P(z_j, t_j), P(z_{j+1}, t_{j+1}))$, hence they will subscribe to tier j (recall also that we have defined $P(z_{p+1}, t_{p+1}) = \infty$, and that $F(P(z_{p+1}, t_{p+1})) = 1$). Also, constraint (6.7) states that the price of a service tier has to be no greater than the utility (value) of this tier to users, since otherwise users will not subscribe even if their budget allows them to do so.

We have the following result.

Lemma 6.1.1 *Let $Z = \{(z_1, t_1), \ldots, (z_p, t_p)\}$ be an optimal solution to MAX-ES-2D. Let $u_j = U(z_j, t_j), j = 1, \ldots, p$. Then, for all j, tier (z_j, t_j) is the point on the indifference curve $U(x, y) = u_j$ that minimizes the cost $C(x, y)$.*

Proof. By contradiction. Assume that in the optimal solution the j-th tier is such that $C(z_j, t_j)$ is not the minimum cost point on the indifference curve $U(x, y) = u_j$. Let (z'_j, t'_j) be such a minimum cost point, and let Z' be the solution derived from Z with (z_j, t_j) replaced by (z'_j, t'_j). Since the utility and price of the j-th tier is not affected by this change, from (6.5) it is clear that $\bar{Q}_{pr}(Z') > \bar{Q}_{pr}(Z)$, contradicting the assumption that Z is an optimal solution. ■

6.2 Approximate Solution to MAX-ES-2D

6.2.1 The Fixed Tier Case

Consider first a special variant of the MAX-ES-2D problem in which the p service tiers are predetermined and part of the input, and not subject to optimization; this variant arises in the case of the uniform and exponential tiering structures that we introduce in Section 6.3. The cost of each tier is completely determined in this case, and for simplicity we let $C_j = C(z_j, t_j), j = 1, \ldots, p$. The price of each tier j is equal to the utility, i.e., $P_j = P(z_j, t_j) = U(z_j, t_j)$, and we let $P_0 = P(x_{min}, y_{min}) = U(x_{min}, y_{min})$. Hence, the provider's surplus for a fixed tier structure Z can be obtained from the following expression:

$$\bar{Q}(Z) = \sum_{j=0}^{p-1} (P_j - C_j)(F(P_{j+1}) - F(P_j)) \tag{6.9}$$

6.2.2 Cost Minimization on an Indifference Curve

Before we tackle the general version of the MAX-ES-2D problem, we note that, because of Lemma 6.1.1, each tier in an optimal solution is the point on an indifference curve with the minimum cost among all points on this curve. Therefore, let us consider the optimization problem of the form:

$$\text{Minimize } \; C(x,y) \; \text{ subject to } \; U(x,y) = u. \tag{6.10}$$

Depending on the form of the cost and utility functions, this problem may be solved exactly or approximately using standard optimization techniques. Here we will only consider cost functions $C(x,y)$ that are linear functions of x and y:

$$C(x,y) \;=\; c_1 x \;+\; c_2 y. \tag{6.11}$$

Assuming Cobb-Douglas utility functions as in (6.3), we may solve for y as a function of x:

$$y \;=\; \left(\frac{u}{x^\alpha}\right)^{1/(1-\alpha)} \tag{6.12}$$

Substituting this value of y into the cost function (6.11), we obtain an expression for the cost that is a function of x only:

$$C(x) \;=\; c_1 x \;+\; c_2 \left(\frac{u}{x^\alpha}\right)^{1/(1-\alpha)}. \tag{6.13}$$

The first and second derivatives of $C(x)$ are:

$$C'(x) \;=\; c_1 + c_2 \, u^{\frac{1}{1-\alpha}} \left(\frac{-\alpha}{1-\alpha}\right) x^{\frac{-1}{1-\alpha}} \tag{6.14}$$

$$C''(x) \;=\; c_2 \, u^{\frac{1}{1-\alpha}} \frac{\alpha}{(1-\alpha)^2} x^{\frac{\alpha-2}{1-\alpha}}. \tag{6.15}$$

If there are no other constraints, we can just let $C'(x) = 0$, and obtain the optimal values:

$$x^\star = u \left(\frac{c_1(1-\alpha)}{c_2\alpha}\right)^{\alpha-1}, \quad y^\star = u \left(\frac{c_1(1-\alpha)}{c_2\alpha}\right)^{\alpha}. \tag{6.16}$$

It is easy to prove that $C''(x^\star) > 0$. Thus, $C(x)$ achieves its minimum value at $x^\star$, hence the original cost function $C(x, y)$ is minimized at $(x^\star, y^\star)$.

Recall, however, that x and y are defined only between respective minimum and maximum values. Consider the above optimization problem under the additional constraints:

$$x_{min} \le x \le x_{max} \qquad y_{min} \le y \le y_{max}. \tag{6.17}$$

It easy to see that when $x < x^\star$, $C'(x) < 0$, and when $x > x^\star$, $C'(x) > 0$. Consequently, whenever the unconstrained minimum point $(x^\star, y^\star)$ from (6.16) lies outside the feasible region defined by constraints (6.17), the minimum point within the feasible region can be obtained as follows:

- if $x^\star > x_{max}$, then $x^\star = x_{max}$, and $y^\star = \left(\frac{u}{x_{max}^{\alpha}}\right)^{\frac{1}{1-\alpha}}$;
- if $x^\star < x_{min}$, then $x^\star = x_{min}$, and $y^\star = \left(\frac{u}{x_{min}^{\alpha}}\right)^{\frac{1}{1-\alpha}}$;
- if $y^\star > y_{max}$, then $x^\star = \left(\frac{u}{y_{max}^{1-\alpha}}\right)^{\frac{1}{\alpha}}$, and $y^\star = y_{max}$; and
- if $y^\star < y_{min}$, then $x^\star = \left(\frac{u}{y_{min}^{1-\alpha}}\right)^{\frac{1}{\alpha}}$, and $y^\star = y_{min}$.

6.2.3 Service Tier Optimization

The most general version of the MAX-ES-2D problem involves the selection of service tiers and their respective prices so as to maximize provider surplus, subject to the constraints (6.6)-(6.8). The utility function $U(x, y)$ provides a relative ranking of service bundles (x, y) in terms of user preference, and the utility of any service tier will lie in the interval $[U_{min}, U_{max}]$, where $U_{min} = U(x_{min}, y_{min})$ and $U_{max} = U(x_{max}, y_{max})$. Therefore, the problem can be logically decomposed into two subproblems:

1. find the indifference curve I_j (i.e., utility value $u_j \in [U_{min}, U_{max}]$) on which each optimal service tier (z_j, t_j) lies and set the price of the tier to u_j; and
2. set tier (z_j, t_j) to the point in indifference curve $I_j, j = 1, \ldots, p$, that minimizes the provider cost $C(z_j, t_j)$.

The second subproblem was addressed in the previous subsection. Next, we develop a dynamic programming solution for the first subproblem.

To this end, we employ a discretization technique. Specifically, we divide the domain $[U_{min}, U_{max}]$ of the utility function $U(x)$ into $K > p$ equal-length sub-intervals, such that the right endpoint U_k of the k-th sub-interval is $U_k = \frac{k(U_{max}-U_{min})}{K}, k = 1, \ldots, K$. We also restrict the tiers to take values only from the discrete set $\{U_k\}$ of indifference curves, rather than the continuous set $[U_{min}, U_{max}]$. Let $\Upsilon(k, l, w)$ denote the optimal value of (6.5) when there are k sub-intervals, l tiers and the l-th tier is set at the indifference curve of utility value $U_w, w \leq k$. Let also $C_w^\star, w = 1, \ldots, K$, denote the minimum cost on the indifference curve of utility U_w. Then, we may write the following recursion:

$$\begin{aligned}
\Upsilon(k,1,w) &= (U_w - C_w^\star)(F(U_k) - F(U_w)), \\
&\quad k = 1, \ldots, K; w = 1, \ldots, k && (6.18) \\
\Upsilon(k,l+1,w) &= \max_{q=l,\ldots,w} \Big\{(U_w - C_w^\star)(F(U_k) - F(U_w)) \\
&\quad + \max_{v=l,\ldots,q} \{\Upsilon(q,l,v)\}\Big\} \\
l = 1, \ldots, p-1; &\quad k = 2, \ldots, K; w = 1, \ldots, k. && (6.19)
\end{aligned}$$

Expression (6.18) can be explained by noting that when there are k sub-intervals and only one tier with a price set to U_w, the customers who subscribe to the service at this price are those with budgets equal to or greater than U_w, or a fraction $(F(U_k) - F(U_w))$ of the total user population. For each subscriber, the provider has a profit of $U_w - C_w^\star$, hence the expected surplus is given by (6.18). Expression (6.19) can be similarly explained. Once $\Upsilon(k, l, w)$ has been computed for all values of k, l, and w, the overall optimal for p tiers and K intervals can be determined as:

$$\max_w \; \Upsilon(K, p, w) \tag{6.20}$$

The overall running time complexity of this dynamic programming algorithm is $O(pK^4)$.

As $K \to \infty$, this discrete version of MAX-ES-2D approaches the original version in which the tiers are continuous variables. We have conducted a large number of experiments (omitted due to space constraints) which indicate that $K = 100$ is sufficient for the dynamic programming algorithm to converge; hence, we use this value in the performance study we present in the next section.

6.3 Numerical Results

In order to evaluate tiering structures for service bundles, we consider an ISP offering a bundle of two services, namely, access speed x and web hosting traffic handled y. The domain of service x is [256 Kbps, 12 Mbps], while the domain of service y is [100 MB, 1 TB]. We consider the following tiering structures in our study:

1. **Optimal:** the set of tiers $Z = \{(z_1, t_1), \ldots, (z_p, t_p)\}$ obtained as a solution to the dynamic programming algorithm (6.18)-(6.20), where $z_i \in$ [256 Kbps, 12 Mbps] and $t_i \in$ [100 MB, 1 TB].

2. **Optimal-rounded:** the set of tiers obtained after rounding the values of each tier $(z_i, t_i) \in Z$ such that z_i is rounded to the nearest multiple of 256 Kbps and t_i is rounded to the nearest multiple of 100 MB.

3. **Uniform-uniform:** the tier structure constructed by (1) obtaining a uniform tiering structure $\{z_1, \ldots, z_p\}$ for service x by spreading the p tiers across the domain [256 Kbps, 12 Mbps], (2) obtaining a uniform structure $\{t_1, \ldots, t_p\}$ for service y by spreading the p tiers across the domain [100 MB, 1 TB], and (3) pairing the tiers of same index in the two sets to form the tiers $\{(z_1, t_1), \ldots, (z_p, t_p)\}$ for the bundle.

4. **Exponential-exponential:** this tier structure is obtained in a similar manner as uniform-uniform, except that the p single-service tiers divide their respective domain into exponential intervals (i.e., intervals that double in length, from left to right).

5. **Uniform-exponential:** the tier structure in which p uniform (respectively, exponential) tiers are obtained for service x (respectively, service y), which are then paired to obtain the p tiers for the service bundle.

6. **Exponential-uniform:** the tiers for service x are exponential and those of service y are uniform.

Note that uniform and exponential tiered structures are similar to those employed by major ISPs (e.g., ADSL tiers of 768 Kbps, 1.5 Mbps, 3 Mbps, 6 Mbps, etc). For the last four tiering solutions, the $p > 1$ service tiers are fixed. Therefore, the provider surplus in this case was obtained from expression (6.9).

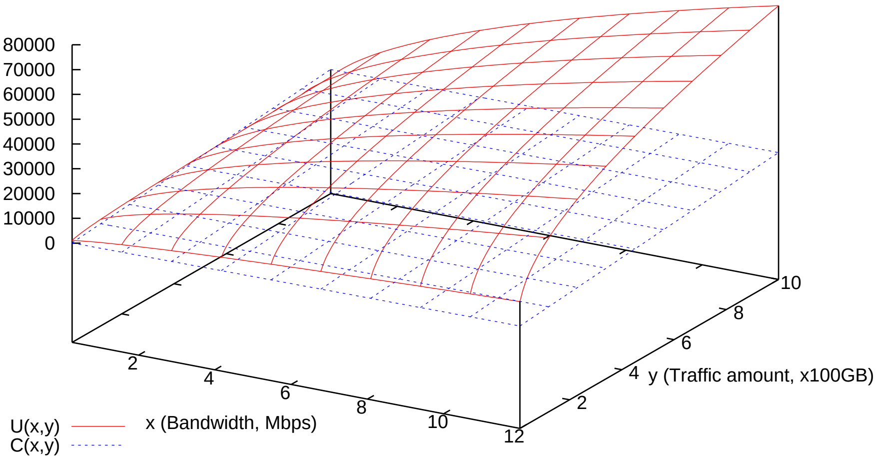


Figure 6.2: Utility and cost functions used for the experiments

We use the Cobb-Douglas utility function in expression (6.3) with parameter $\alpha = 0.6$, and a linear cost function as in expression (6.11), with $c_1 = 0.1$ and $c_2 = 0.01$; these values for c_1 and c_2 were selected so that neither term of the cost function dominates across the domains of services x and y. Plots of the utility and cost functions are shown in Fig. 6.2.

In order to study the effect of the distribution of user budgets, we consider three distinct distributions in the domain $[U_{min}, U_{max}]$:

- a *decreasing* distribution, $f(B) = -\frac{2B}{(U_{max}-U_{min})^2} + \frac{2U_{max}}{(U_{max}-U_{min})^2}$, in which the mass of the distribution is concentrated at lower budget values (less affluent population),
- a *uniform* distribution with PDF $f(B) = \frac{1}{U_{max}-U_{min}}$, and
- an *increasing* distribution, $f(B) = \frac{2B}{(U_{max}-U_{min})^2} - \frac{2U_{min}}{(U_{max}-U_{min})^2}$, in which the mass of the distribution is concentrated at higher budget values (more affluent population).

Figures 6.3, 6.4, and 6.5 plot the expected provider surplus for the uniform, increasing and decreasing, respectively, distribution of user budgets. Each figure shows six

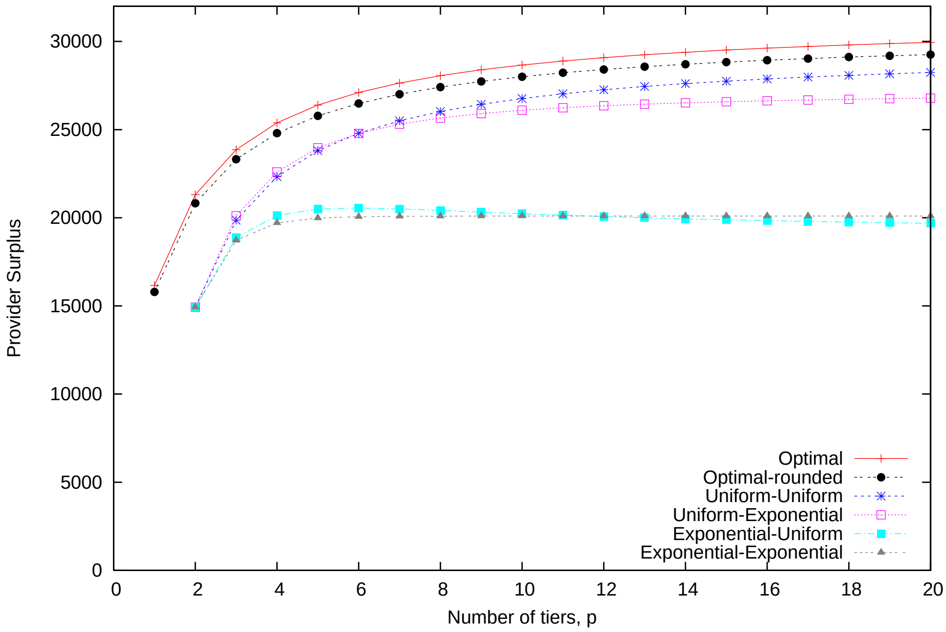


Figure 6.3: Tiered structure comparison, uniform budget distribution

curves, corresponding to the six tiered structures above. A first observation is that, for a given tiered structure and a given number of tiers, the expected provider surplus depends directly on the distribution of user budgets. Specifically, the provider surplus increases from Fig. 6.5 (decreasing distribution) to Fig. 6.3 (uniform distribution) to Fig. 6.4 (increasing distribution). This result is directly due to the fact that the average user budget is lowest under the decreasing distribution and highest under the increasing distribution.

We also observe that the optimal and optimal-rounded structures outperform the other four fixed-tier structures. The optimal-rounded curves lie a little lower than the corresponding optimal curves, as a result of rounding in two dimensions. More importantly, structures which include exponential tiering of at least one service are the worst performers in terms of provider profits. This behavior demonstrates that exponential tiers currently favored by major ISPs are far from optimal. Overall, these results provide a strong indication that the optimization methodology we developed in this chapter represents a valuable tool for service providers.

6.4 Concluding Remarks

In this chapter we have investigated tiered structures for bundles of network services with the objective of maximizing provider profits under user constraints. We have developed an efficient dynamic programming algorithm for determining jointly the service tiers and their prices. Although we only considered bundles of two services, our work may be extended to bundles of more than two services.

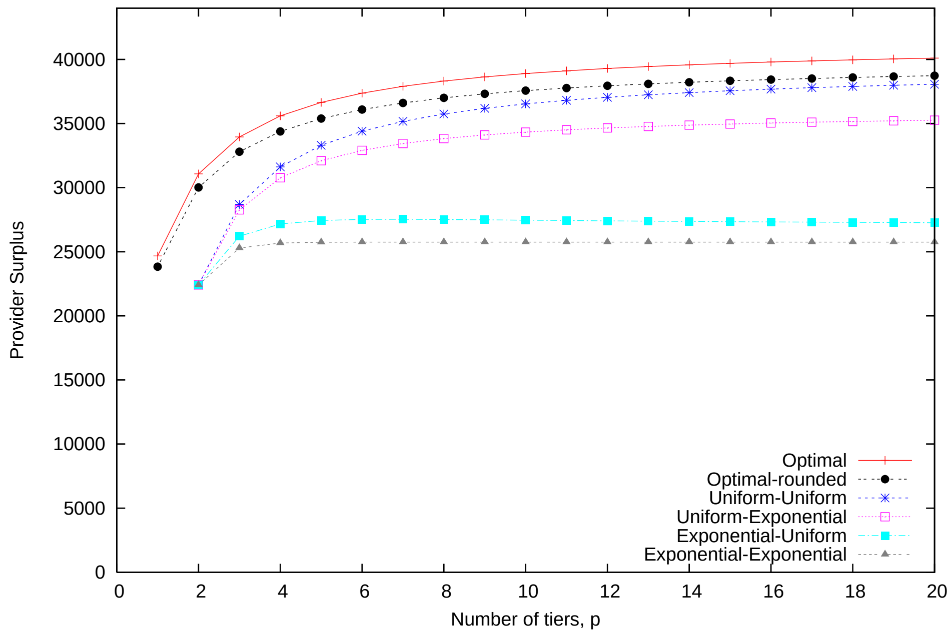


Figure 6.4: Tiered structure comparison, increasing budget distribution

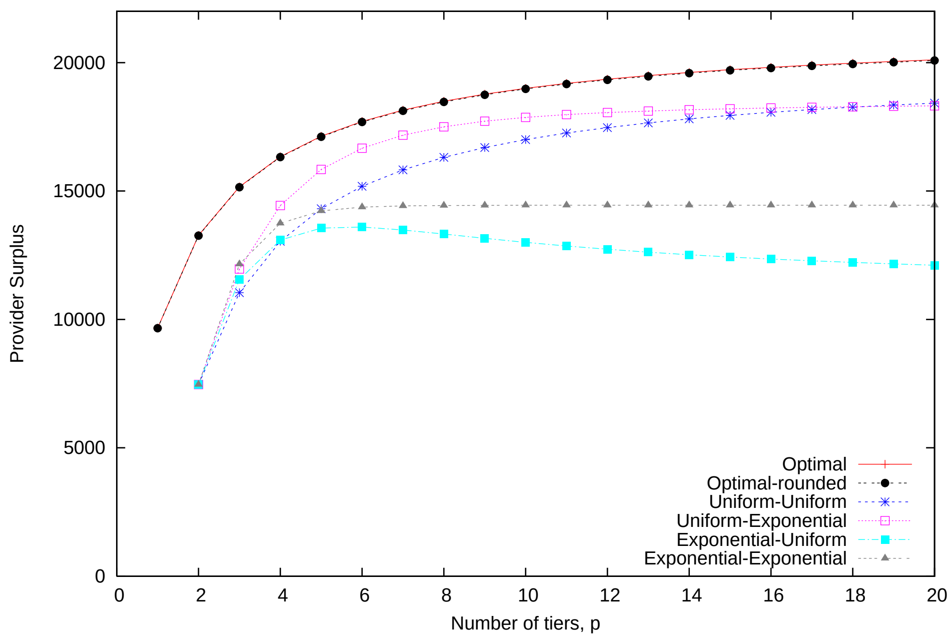


Figure 6.5: Tiered structure comparison, decreasing budget distribution

Chapter 7

Summary and Future Work

In this thesis, we have discussed economic issues in tiered network services. Our work can be summarized as follows:

Firstly, we discuss the service tier location method (i.e., sizing the network) for the network with discrete service demand requests. We proposed dynamic programming algorithms on optimal selection of the service tiers by considering the interests from the customers, service providers, or the whole society.

Secondly, we discuss the service tier location method for the network with stochastic service demand requests. We consider the problem as a typical nonlinear programming model. We propose a (near-)optimal solution by quantizing the bandwidth into small intervals and choose the tiers from the interval bounders. We then use tools from Nash bargaining in game theory to formulate the pricing setting in each tier.

Thirdly, we consider the Internet broadband access as an elastic service whose value varies across segments of the user population. We show that introducing multiple tiers of service can be an effective market sgementation strategy that lead to an increase of profits for the ISP. We also develop an efficient dynamic programming algorithm for the problem of determining optimally both the service tiers and their prices.

Fourthly, we consider the service bundling problem, which means the provider combine several services together and sold them as a single package with affordable price. We consider the problem of determining optimal tiering structures for service bundles using tools from economics and utility theory. Our work provides insight into the selection and pricing of Internet tiered services.

All the above methods are proved - through simulation - to be more effective when comparing with the existing methods employed by major ISPs for choosing and pricing tiered servcesx. We believe these methods can help the ISPs to design and manage tiered services more effectively.

7.1 Directions for Future Work

Our work can be extended in several directions.

1. **Extensions to multi-provider market.** In this thesis, the service tier location methods assume that there exist only one service provider in the market. We would like to relax this assumption by consider multi-provider market scenarios. By doing this we hope to employ knowledge from game theory to simulate the competition among different service providers.

2. **Solving multi-dimensional directional p-Median problem.** Directional P-Median problem can be used to model many telecom applications, including tiered service management, dimensioning of MPLS tunnels for VPNs, packet fair queueing, task scheduling in multiprocessor, cluster or grid systems, etc. The one-dimensional directional p-Median problem (e.g., n demand points and p service points) can be solved by using dynamic programming algorithm with $O(np)$ time complexity. The multi-dimensional directional p-Median problem, however, is proved to be NP-Complete [17]. Currently methods to solve such problem, like TB heuristic [35] and DH heuristic [16], are either too time-consuming or not accurate enough (i.e., the solution is not near-optimal). We plan to develop a new method to solve the multi-dimensional directional p-median problem with low time complexity and high accuracy.

3. **Adavanced service bundling method.** In the service bundling method we proposed in Chapter 6, we assume the service provider knows a priori the customer indifference curves. This may be impractical in reality because sometimes it may be difficult to compare the customers' perference for different services/products. In addition, the provider may bundle three or more services/products together; the multi-dimensional indifference curve is more diffcult to obtain. To address such challenges, we plan to apply the knowledge on multi-dimensional directional p-Median problem and microeconomic theory to develop an efficient method to solve service bundling problem in general cases.

4. **In depth experimental investigation with various utility and cost functions.** Our methods in the thesis apply to any kinds of utility functions and cost functions. However, the effects of our methods may vary with the definition of the utility or cost functions. We would like to look for more realistic utility and cost functions, and perform in depth study on the effects of our methods on these functions.

Bibliography

[1] A. Aggarwal and J. Park. Notes on searching in multidimensional monotone arrays. *Symposium on Foundations of Computer Science*, 0:497–512, 1988.

[2] 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.

[3] H. L. Ahuja. *Modern Microeconomics: Theory and Applications*. S. Chand and Company Ltd., 2004.

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

[5] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2nd edition, 2004.

[6] G. R. Bitran and J. C. Ferrer. On pricing and composition of bundles. *Production and Operations Management*, 16(1):93–108, February 2007.

[7] K. E. Case and R. C. Fair. *Principles of Economics*. Pearson Education, 7th edition, 2004.

[8] G. Cornuejols, G.L. Nemhauser, and L.A. Wolsey. The uncapacitated facility location problem. In *P.B. Mirchandani and R.L. Francis (Editors), Discrete Location Theory*, pages 119–171. Wiley, 1998.

[9] Rudra Dutta, Ahmed E. Kamal, and George N. Rouskas. *Traffic Grooming for Optical Networks: Foundations, Techniques and Frontiers*. Springer Publishing Company, Incorporated, 2008.

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

[11] Cooperative Association for Internet Data Analysis (CAIDA), January 2008.

[12] C. Holahan. Time warners net metering precedent. *Business Week*, June 2008.

[13] C. Holahan. Time warners pricing paradox. *Business Week*, January 2008.

[14] C. Huang, J. Li, and K. W. Ross. Can internet video-on-demand be profitable? In *Proceedings of ACM SIGCOMM*, pages 133–144, 2007.

[15] Laura E. Jackson, Rouskas George N., and Stallmann Matthias F.M. The directional p-median problem: Definition, complexity, and algorithms. *European Journal of Operational Research*, 179(3):1097–1108, June 2007.

[16] Laura E. Jackson, George N. Rouskas, and Matthias F.M. Stallmann. The directional p-median problem: Definition, complexity, and algorithms. *European Journal of Operational Research*, 179(3):1097–1108, June 2007.

[17] Laura E. Jackson and Matthias F. Stallmann. Directional p-median in two dimensions is np-complete. Tr-2004-10, Department of Computer Science, North Carolina State University, 2004.

[18] Ehud Kalai. Solutions to the bargaining problem. Discussion Papers 556, Northwestern University, Center for Mathematical Studies in Economics and Management Science, March 1983.

[19] N. Kausar, B. Briscoe, and J. Crowcroft. A charging model for sessions on the internet. In *European Conference on Multimedia Applications, Services and Techniques*, pages 246–261, 1999.

[20] Qian Lv and George N. Rouskas. On optimal sizing of tiered network services. In *Proceedings of INFOCOM*, April 2008.

[21] Qian Lv and George N. Rouskas. An economic model for pricing tiered network services. In *Proceedings of ICC*, June 2009.

[22] Qian Lv and George N. Rouskas. An economic model for pricing tiered network services. *Annals of Telecommunications*, 2009.

[23] Qian Lv and George N. Rouskas. Internet service tiering as a market segmentation strategy. In *Proceedings of GLOBECOM*, November 2009.

[24] Qian Lv and George N. Rouskas. On optimal tiered structures for network service bundles. *Submitted*, 2009.

[25] L. Moutinho and A. Meidan. Quantitative methods in marketing. In *M. J. Baker (Editor), The Marketing Book*, pages 197–244. Butterworth-Heinemann, 2002.

[26] J. F. Nash. The bargaining problem. *Econometrica*, 18:155–162, 1950.

[27] J. F. Nash. Two-person cooperative games. *Econometrica*, 21:128–140, 1950.

[28] Martin J. Osborne and Ariel Rubinstein. *Bargaining and Markets*. Academic Press, 1990.

[29] Arthur Cecil Pigou and Nahid Aslanbeigui. *The Economics of Welfare*. Business & Economics, 1932.

[30] George N. Rouskas. *Internet Tiered Services: Theory, Economics, and Quality of Service*. Springer Publishing Company, Incorporated, 2009.

[31] George. N. Rouskas and N. Baradwaj. A framework for tiered service in MPLS networks. In *Proceedings of IEEE INFOCOM 2007*, May 2007.

[32] George N. Rouskas and Nikhil Baradwaj. Tdm emulation in packet-switched networks. In *ICC*, pages 1911–1916, 2007.

[33] Chipman John S. and James C. Moore. The new welfare economics 1939-1974. *International Economic Review*, 19(3):547–584, 1978.

[34] S. Shenker. Fundamental design issues for the future internet. *IEEE Journal on Selected Areas in Communications*, 13(7):1176–1188, September 1995.

[35] M. B. Teitz and P. Bart. Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations Research*, 14:955–961, 1968.

[36] H. Wang, H. Xie, L.Qiu, A. Silberschatz, and Y.R. Yang. Optimal isp subscription for internet multi-homing: Algorithm design and implication analysis. In *Proceedings of IEEE INFOCOM*, pages 2360–2371, 2005.

[37] A. Weinstein. *Market Segmentation.* Probus Publishing Co., 1993.

[38] Randall Wright. Bargaining theory. Class notes, Department of Economics, University of Pennsylvania.