EFFICIENT CAPACITY PRICING OF THE INTERNET SERVICES

CHANG-HO YOON

Department of Economics, Korea University
Seoul 136-701, Korea
chyoon@korea.ac.kr

YOUNG-WOONG SONG

SK Research Institute for SUPEX Management
Seoul 110-110, Korea
ywsong@sktelecom.com

B YOUNG HEON JUN

Department of Economics, Korea University
Seoul 136-701, Korea
bhjun@korea.ac.kr

Accepted February 2005

Abstract

The paper examines the possibilities to improve efficiency in Internet pricing by introducing pre-purchase contract. One can regard pre-purchase market as a device for providing guaranteed services and as an alternative to smart market that can implement expected capacity pricing in an efficient manner. We find that the pre-purchase market tends to discriminate against the consumers who are less certain about their demands. We provide a condition under which the discriminatory effect is overwhelmed by the market force, which discourages the consumers with lower value by high premium. We also suggest a solution to the discriminatory effect.

Keywords: Network congestion, Pre-purchase, Edge pricing, Smart market, Internet pricing

JEL Classification: D60, D81, L86

I. Introduction

As the social demand for bandwidth ever increases over time, congestion in computer networks seems inevitable. Despite extensive supply of network infrastructure and increasing availability of bandwidth, entrepreneurial exploration of bandwidth-intensive applications and services in content industries seems to put no upper bound on the use of bandwidth. Potential scarcity of future bandwidth seems to have necessitated in-depth analysis of pricing policy in
modern computer networks and reconsideration of the flat rate pricing under over-provisioning policy (McKnight and Boroumand, 2000). Various concepts of efficiency that incorporate not only economic efficiency but architectural network efficiency and administrative costs have been examined.

The first monumental market mechanism to resolve network congestion was introduced by MacKie-Mason and Varian (1995) who then theoretically applied incentive-compatible Vickrey auction to determine priorities among the packets of data that are waiting at each node of their path toward final destination. Although the smart market mechanism is ex post economically efficient, feasibility of implementation has become more important concern to many researchers. In addition to administrative costs of implementing a numerous number of bidding mechanisms at each node of long and complicated configuration of networks that are connected by routers, there would be the costs of calculating the utility loss of delay that are caused by retransmission of packets of losing bidders, which would normally be prohibitive. The problems related to valuation of sequences of packets reveal another aspect of technical difficulties associated with implementation of smart market mechanism (see for example, Crémer and Hariton, 1999 in this direction).

Doubts about feasibility of smart market mechanism and inaccessibility of marginal congestion costs lead to reshaping the research agenda toward technically more tractable mechanism that however retains some elements of usage sensitivity and localized pricing policy (see for example, Shenker, Clark, Estrin and Herzog, 1996, and Clark, 1997). The so called edge pricing and expected capacity pricing are both implemented locally at the access point or the edge of the ISP’s network where the user’s packets enter. Once the source and the final destination are known, the entire computation of charges is performed at the access point, and is based on the expected capacity requirement and expected congestion costs along the expected path of packets between the source and destination. It performs like time-of-day telephone pricing that depends only on expectations about the current congestion costs, and is not sensitive to instantaneous traffic condition. The shift of emphasis from the per-packet charge to the user cost of purchasing the required capacity for transmission of their information rekindle interests in the flat rate pricing since it is essentially defined as a peak rate and can be modified in a flexible manner whenever a user needs more than the capacity that was previously purchased. It may in fact take various forms of mixtures of capacity-based prices and usage-based prices. The flat rate pricing system offers a starting point to develop higher quality of service in general.

Although this possibility was already examined in the edge pricing literature, the efficiency comparison of various modes of capacity pricing was not on the urgent research agenda. Especially the theoretical possibility of enhancing efficiency by introducing a contract to pre-purchase expected required capacity was not noticed in the literature. The new pricing policy may incorporate intertemporal variation of expected capacity pricing that reflects users’ valuation more accurately in pricing of network capacity. Under the flat rate system, users have equal chance to get allocated the scarce capacity independently of their willingness to pay. Under pre-purchase contract only those who are willing to pay high price get the greater chance to use the capacity when congestion occurs. This treatment opens the possibility of increasing allocative efficiency.

However, pre-purchase system is not always more efficient than the flat rate system when there is no appropriate mechanism to distinguish those frequent users who have nearly
constant demand across various states of natures from non-frequent users whose demand fluctuates across various states of nature. It may be the case that non-frequent users might have valuable messages to send but the utility value of messages for them may fall short of the equilibrium premium price. In fact the network operator would sell the greater amount of capacity than is actually used in the market clearing state knowing that only the portion of the contracted capacity for non-frequent users will actually be used. The downward pressure in the pre-purchase contract price induces frequent users to contract more of the capacity for their use while the resulting effective capacity price that the non-frequent user pays in equilibrium turns out to be much greater than the price that the network provider receives at the market equilibrium.

The efficiency gain from the use of pre-purchase contract requires the additional assumption that expected capacity cost of non-frequent users must be less than the average surplus value of frequent users. If this assumption is not satisfied, it is always possible for some of non-frequent users to form coalition and share the capacity by distributing the cost among themselves. It is certainly socially more efficient to allocate the capacity to them than to frequent users who send messages only of low value. One way to discriminate them is to give refund to those whose usage time is low. This requires monitoring of usage and technically more demanding. If however it is possible to monitor each consumer’s usage, it is always possible to make an equilibrium with pre-purchase contract ex ante efficient.

The purpose of this paper is to investigate efficiency of capacity pricing. Starting from the flat rate system, the paper presents a theoretical model to examine the possibility of improving efficiency with minimal additional technical requirements. The paper does not focus on how to calculate expected capacity and congestion costs that are associated with the entire transmission path. Neither the paper deals with formation of users’ expectations on the extent of congestion. Rather it addresses the question regarding the efficient use of existing scarce capacity among users who have differing valuation of transmission of information through Internet. The paper begins with a basic model of congestion and efficiency analysis of the flat rate system and pre-purchase contract. Comparison with other market mechanism to resolve congestion is presented. The concluding comments on technical consideration that are necessary to implement pre-purchase contract briefly follows, and related research agenda on quality of service provision will be briefly discussed.

II. Model

We begin with a basic model where the expected capacity required between the source and the final destination is already decided and normalized to be one unit, therefore consider a very simple network where just one representative operator (ISP) provides Internet access service. We assume that traffic generated from the source will be assumed to be delivered to the final destination via the single node. In this sense network congestion can be defined to be a state where the capability of the node (capacity) doesn’t meet incoming traffic’s demand any more.

Internet service user may have two-dimensional requirement for Internet services. One dimension is for ‘Quality of Service’ requirement and the other is for ‘Quantity of Service’ requirement. Though there are seemingly differences between two kinds of QoS requirements, they are intrinsically the same in that the further we go up in either scale, the more capacity
ISP needs to provide. In another perspective, we can say that ISP has time-based capacity conditional on the same Quality of Service requirement and bandwidth-based capacity given the same quantity requirement. From now on we assume that the two-dimensional requirement can be integrated, so that consumers' requirements (demand) and ISP's constrained capacity (supply) is denoted by a certain unit (e.g. bps \times number of packets or length of period).

There are two states of nature, $s_1$ and $s_2$, state $s_i$ occurring with probability $\mu_i$, and two types of consumers. Those with certain demand (type c) will have demands in both states with probability 1, and those with uncertain demands (type u) will have demands with probability $\theta_1$ in state $s_1$ and with probability $\theta_2$ ($>\theta_1$) in state $s_2$. For simplicity we assume that each consumer demands only one unit,\(^1\) and that there are a continuum of consumers, measure $a$ of type c and measure $b$ of type u. Without loss of generality we normalize $b$ to 1. The value that consumers derive from consumption of internet service (normalized by a defined unit) is assumed to be distributed on the interval $[v, v]$ with distribution functions $F$ for type c and $G$ for type u. The value $v$ is assumed to be known to every consumer in advance, although its realization is uncertain for type u consumers. The type and value of consumers are assumed to be private information, although the distribution functions $F$ and $G$ are common knowledge.

Under these assumptions market demand will be $a + \theta_1$ in state $s_1$ and $a + \theta_2$ in $s_2$. Although individual type u consumers have uncertain demands, there is no aggregate uncertainty in the market demand due to the assumption of “large number” (continuum) of consumers. We denote the total capacity (during a certain period) by $k$ and assume that $a + \theta_1 < k < a + \theta_2$. Hence not all demands are satisfied in state $s_2$. When there is no mechanism to curtail the demand in state $s_2$, rationing should take place. This is what happens under the flat rate system. Unless explicitly stated otherwise, we assume that consumers and suppliers are risk neutral.

III. Flat Rate System vs. Pre-purchase Market

1. Flat Rate System

Under the flat rate system everyone is treated equally. When congestion occurs the first comer gets served first. Since each consumer has equal chance to be a first comer, each consumer gets served with probability $k/(a + \theta_2)$ in $s_2$, the congestion state. In order to illustrate what happens we consider an example in which $a = 1$, $\theta_1 = 0.2$, $\theta_2 = 0.8$, $k = 1.5$, $\mu_1 = \mu_2 = 1/2$, and the value of both types are uniformly distributed on $[1, 2]$. In $s_1$, the non-congestion state, market demand is less than the capacity and everyone who happens to have demand gets the service. On the other hand, in $s_2$, the congestion state, the consumers who have demand get served with probability 5/6.\(^2\) The ex ante total surplus is

---

\(^1\) Therefore, consumers in our model are homogeneous in Quality and Quantity of Service requirements. If one wants to include heterogeneous consumer (in terms of Quality of Service requirement) into the model, he or she can do so by introducing more general demand function.

\(^2\) We assume that the (fixed) price is so low that everyone is subscribed. Alternatively, we focus on the consumers who are subscribed under the current price.
2. Pre-purchase Market

We consider the performance of a pre-purchase market under the same environment considered in the previous example. In the pre-purchase market consumers purchase the right to use the service in congestion state for a premium. In effect they buy a kind of contingent contract that will insure consumption in congestion state. If congestion occurs, the suppliers allocate the capacity first to the consumers who paid the premium, and then to those who did not if there is any remaining capacity. Hence, in order for the pre-purchase market to function properly we need to have a device in the network that will distinguish the consumers who paid the premium from those who did not.\(^1\) To have such a device could be costly, but certainly less costly than the device required for the smart market.

We assume that the suppliers set the premium in such a way that all the capacity is sold in the pre-purchase market. This would be the case if the pre-purchase market were competitive. It also looks plausible to allocate scarce resources to those who are willing to pay the most. Hence the government may want to enforce such an outcome when the suppliers are regulated. As in the previous subsection we maintain the assumption that \(a=1, \theta_1=0.2, \theta_2=0.8, k=1.5, \mu_1=\mu_2=1/2\), and that the value of both types are uniformly distributed on \([1, 2]\).

First, we calculate the level of the premium for the pre-purchase contract. That is, we will find out the price which will equate the demand and the given supply. The ex ante value of the contract to a type \(c\) consumer is \(v\). Since there is congestion with probability \(1/2\), and since \(v/2\) is uniformly distributed on \([1/2, 1]\) with density 2, the demand of type \(c\) consumers is

\[
D_c(p) = \int_{p}^{1} dF^*(v/2) = 2(1-p), \text{ for } 1/2 \leq p \leq 1,^4
\]

where \(F^*\) is the distribution function of \(v/2\) of type \(c\) consumers. Similarly, the demand of type \(u\) consumers is

\[
D_u(p) = \int_{p}^{0.8} dG^*(0.4v) = 2.5(0.8-p), \text{ for } 0.4 \leq p \leq 0.8,^5
\]

where \(G^*\) is the distribution function of \(0.4v\) of type \(u\) consumers. The price is determined at the level where demand equals supply. Since only 80% of type \(u\) consumers will actually need the service, the equilibrium requires

\[
D_c(p) + 0.8D_u(p) = 1.5
\]

From (2), (3), and (4), we get \(p = 0.525\). Hence, 0.95 of type \(c\) consumers and 0.6875 of type

\[\text{\footnotesize注}^1\text{One such device is a dedicated line.}\]

\[\text{\footnotesize注}^4\text{D}_c(p)=1, \text{ for } p < 1/2.}\]

\[\text{\footnotesize注}^5\text{D}_u(p)=1, \text{ for } p < 0.4.}\]
u consumers buy the contract. Total surplus is
\[
TS^p = \frac{1}{2} \left( \int_1^2 v \, dv + 0.2 \int_1^2 v \, dv \right) + \frac{1}{2} \left( \int_1^{1.05} v \, dv + 0.8 \int_{1.3125} v \, dv \right) = 2.07984
\] (5)
Comparing (1) with (5), one can see that the pre-purchase contract yields higher total surplus.

Under the fixed rate system everyone gets equal chance to be served, whether he/she values the service high or low. Under the pre-purchase contract, on the other hand, only those who value the service high are served. Then, is it the case that the pre-purchase system always yields better outcome than the flat rate system? The answer is no as the following example shows.

3. An Example

Suppose that \( \mu_1 = 0 \) (only congestion state occurs), \( a = b = 1 \), \( \theta_1 = 0 \), \( \theta_2 = 1/2 \), \( k = 1 + \varepsilon \), where \( \varepsilon \approx 0 \). Suppose further that all type c consumers have the same value, \( v = 1 \), and that all type u consumers have the same value, \( v = 1.8 \). Under the flat rate system everyone has the same chance to be served, and the total surplus is
\[
TS^f = \{(1 + \varepsilon)/1.5\}(1 + 0.5 \cdot 1.8) \approx 1.26667
\]
Under the pre-purchase system type c consumers would pay up to 1, whereas type u consumers would pay up to 0.9. Premium will be 0.9, and all of type c consumers as well as 2\( \varepsilon \) of type u consumers will be served. Total surplus will be
\[
TS^p = 1 + 1.8\varepsilon \approx 1
\]
Clearly the flat rate system performs better. The reason is that the pre-purchase system in the current form favors the consumers who are willing to pay more, regardless of the capacity they require. In this example, the suppliers need 1 unit of capacity to serve one type c consumer, while they need only 1/2 unit of capacity to serve one type u consumer. Since the suppliers cannot distinguish type u consumers from type c consumers, the pricing of pre-purchase market works against type u consumers who is actually less likely to claim the service, hence requires less capacity. This example gives us an idea about how to improve on the pre-purchase system, which we will consider in section IV.

4. General Comparison of the Two Systems

We can now compare the total surpluses obtained under the two systems. The total surplus under the flat rate system is
\[
TS^f = \mu_1 \left( a \int_\chi v \, dF(v) + \int_\chi \theta_1 v \, dG(v) \right) + \mu_2 \rho \left( a \int_\chi v \, dF(v) + \int_\chi \theta_2 v \, dG(v) \right), \quad (1')
\]
where \( \rho = k/(a + \theta_2) \) is the probability that each consumer gets the service. The total surplus under the pre-purchase system is
\[
TS^p = \mu_1 \left( a \int_\chi v \, dF(v) + \int_\chi \theta_1 v \, dG(v) \right) + \mu_2 \left( a \int_{v_c} v \, dF(v) + \int_{v_u} \theta_2 v \, dG(v) \right), \quad (5')
\]
where \( v_c(v_u) \) is the value of the type c (type u) consumer just willing to pay the premium p. v_c
and $v_u$ satisfy the following conditions:

\[
\mu_2 v_c = \mu_2 \theta_2 v_u = p, \quad (6)
\]

\[
D_c(p) = a(1 - F(v_c)), \quad \text{and} \quad (7)
\]

\[
D_u(p) = 1 - G(v_u). \quad (8)
\]

The inverse demand functions for the premium service are derived from (6)-(8), respectively as

\[
P_c(q_c) = \mu_2 F^{-1}(1 - q_c/a), \quad \text{and} \quad (9)
\]

\[
P_u(q_u) = \mu_2 \theta_2 G^{-1}(1 - q_u). \quad (10)
\]

The total surplus derived from the premium service is

\[
T_{Sp} = \int_0^{\infty} P_c(q) dq + \int_0^{\infty} P_u(q) dq \quad (11)
\]

which is the same as the second term of $T_{Sp}$ in (5'). Using the inverse functions we can also rewrite the second term of $T_{Sp}$ in (1') as

\[
T_{Sp} = \rho \left( \int_0^{\infty} P_c(q) dq + \int_0^{\infty} P_u(q) dq \right). \quad (12)
\]

We can now provide a sufficient condition for the pre-purchase system to perform better than the flat rate system. Denote the equilibrium price in the pre-purchase market by $p^*$. $p^*$ is defined by

\[
D_c(p^*) + \theta_2 D_u(p^*) = k. \quad (4')
\]

Define $q_c^* = D_c(p^*)$ and $q_u^* = D_u(p^*)$. Then one can prove the following.

**Theorem.** If $\int_0^{\infty} P_c(q) dq / q_c^* \geq p^*/\theta_2$, then the pre-purchase system yields higher total surplus than the flat rate system.

Before we present the proof, let us examine the meaning of the theorem. The LHS of the inequality is the average value assigned to the premium service by the type $c$ consumers who actually purchase the service. The RHS is the effective price of the premium service faced by type $u$ consumers. Since type $u$ consumers has to pay $p^*$ for the service which they will use only with probability $\theta_2$, the effective price of the service is $p^*/\theta_2$ for them. This is the reason why the pre-purchase system may yield poor performance as illustrated in the example of the previous subsection. One can improve upon the pre-purchase system by choosing one type $u$ consumer who was eliminated in the pre-purchase market and let him/her get the service instead of one type $c$ consumer. If we use the flat rate system, this can be done but only through random selection. The theorem says that even if we choose the best of type $u$ consumer who was eliminated in the pre-purchase market (whose value of the service cannot exceed $p^*/\theta_2$), the efficiency cannot be enhanced if the type $c$ consumers who are in the market value the service more than $p^*/\theta_2$ on average.
Proof of Theorem: From (11) and (12) one can see that \( T \text{Sp} \geq T \text{Sp}' \) if and only if

\[
\frac{a + \theta_2}{k} \left( \int_0^{q^c} P_c(q_c) dq_c + \int_0^{q^u} P_u(q_u) dq_u \right) \geq \int_0^{q^c} P_c(q_c) dq_c + \int_0^{q_u} P_u(q_u) dq_u.
\]

\[
\Leftrightarrow \frac{a + \theta_2 - k}{k} \left( \int_0^{q^c} P_c(q_c) dq_c + \int_0^{q^u} P_u(q_u) dq_u \right) \geq \int_0^{q^c} P_c(q_c) dq_c + \int_0^{q^u} P_u(q_u) dq_u.
\]

\[
\Leftrightarrow [(a - q^*_c) + \theta_2(1 - q^*_u)] \left( \frac{q^c_e}{k} \int_0^{q^c} P_c(q_c) dq_c + (1 - \frac{q^u_e}{k}) \int_0^{q^u} P_u(q_u) dq_u \right)
\]

\[
\geq \int_0^{q^c} P_c(q_c) dq_c + \int_0^{q^u} P_u(q_u) dq_u.
\]

Denote the LHS and RHS of the last inequality by \( A^* \) and \( A \) respectively, then

\[
A \leq \int_0^{q^c} p^* dq_c + \int_0^{q^u} p^* dq_u = (a - q^*_c)p^* + \theta_2(1 - q^*_u)p^*/\theta_2.
\]

\[
< [(a - q^*_c) + \theta_2(1 - q^*_u)]p^*/\theta_2. \quad (13)
\]

Since \( \int_0^{q^c} P_c(q_c) dq_c \geq q^*_c p^* \), if it is true that \( \int_0^{q^u} P_u(q) dq/\theta_2 \geq p^*/\theta_2 \), then we have

\[
A^* \geq [(a - q^*_c) + \theta_2(1 - q^*_u)] p^*/\theta_2. \quad (14)
\]

From (13) and (14), we get the desired inequality. \( \text{(Q.E.D.)} \)

IV. **Smart Market vs. the Pre-purchase Market**

Basically, the model considered in this paper is about capacity-based pricing while smart market is about packet-based pricing, so that there seems to be no general way to compare the two pricing systems. But so far as the quantity of demand and supply are normalized and consumers have homogeneous requirements for Quality of Service as we assumed before,\(^6\) the two pricing systems can be compared with each other.

The two markets allocate the capacity in the same way when there is no congestion (state \( s_1 \)). They allocate the capacity differently when there is congestion (state \( s_2 \)). The difference is shown in Figure 1. In the smart market the capacity is allocated to the consumers who are willing to pay the market clearing price \( p_2 \). \( k_c \) of type \( c \) consumers and \( k_u = k - k_c \) of type \( u \) consumers get to use the capacity. Smart market is ex post efficient, and hence is ex ante efficient in a risk-neutral environment.

On the other hand, the capacity is allocated to the consumers who are willing to pay the premium \( p^* \) ex ante. \( k^*_c \) of type \( c \) consumers and \( k^*_u = (k - k^*_c)/\theta_2 \) of type \( u \) consumers are willing to pay \( p^* \) ex ante. Type \( u \) consumers, however, get the chance to consume (one unit of the capacity) only with probability \( \theta_2 \). Hence, they are effectively paying \( p^*/\theta_2 \) as the premium.

---

\(^6\) This implies that the quantity of demand in our model is a linear function of the number of packets that are the quantity of demand in smart market.
Out of \( k_u^* \) type \( u \) consumers only \( \theta_2 k_u^* \) actually get to use the capacity ex post. Since there is price discrimination as a result, the allocation is not efficient in the pre-purchase market. Is it possible to improve the pre-purchase market? One simple way is to give refunds to those who did not use the service. Then type \( u \) consumers get refunds with probability \( 1 - \theta_2 \), effectively paying \( p^* \). Since everyone pays the same price, the capacity allocation will be the same as in the smart market.

Introduction of refund requires additional technology. The supplier now needs to trace the usage of the consumers who paid the premium. However, this technology would be less demanding than that required for the smart market. It is not hard to construct examples in which the pre-purchase market is ex ante more efficient than the smart market when the consumers are risk averse.

V. Conclusion

Since the smart market was proposed as a solution to the problem of congestion in the Internet, not much research has been conducted on the Internet pricing. It is partly because smart market provides the ultimate answer one can expect at least in risk neutral environment, and partly because Internet congestion is less severe due to technological improvement in data transfer. However, the technological burden that the smart market imposes is not trivial, and it seems natural to ask what alternatives we have to avoid unwanted interruption of data transfer. This becomes more and more important as people want to transfer larger and larger data containing voices and images. New demands are created as people think of new data that they want to transfer. One never knows when the demands catch up the speed of network expansion.

In this paper we investigated the pre-purchase market as an alternative to the smart market and as a device for providing guaranteed services in ex ante perspective. One merit of

\[ p^* \text{ will be the same as } \mu_3 p_2. \]
the pre-purchase market is to allow the user to plan for the future while handling architectural and structural issues. We found that the pre-purchase market tends to discriminate against the consumers who are less certain about their demands. We provided a condition under which the discriminatory effect is overwhelmed by the market force, which discourages the consumers with lower value by high premium. If we regard a local network operator (ISP) as a final user of upstream network, the analysis to more complicated networks can easily be extended.

There are several protocols that meet various kinds of consumer needs for QoS technically, such as RSVP or ST-II protocol. As the specifications of these protocols have not yet been fully developed to implement complicated networks, our framework for guaranteed service may require some more time to be globally implemented. However, confining our focus to local network, we can consider pre-purchase market as an efficient solution to network congestion with least technical difficulties.

REFERENCES


Crémer J. and Hariton C. (1999), The Pricing of Critical Applications in the Internet, Mimeo.


