<table>
<thead>
<tr>
<th>Title</th>
<th>Efficient Allocation of the Internet Resources and Its Pricing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Takashima, Makoto</td>
</tr>
<tr>
<td>Citation</td>
<td>経済学部研究年報, 15, pp.1-32; 1999</td>
</tr>
<tr>
<td>Issue Date</td>
<td>1999-03-25</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10069/26198">http://hdl.handle.net/10069/26198</a></td>
</tr>
</tbody>
</table>

NAOSITE: Nagasaki University’s Academic Output SITE
http://naosite.lb.nagasaki-u.ac.jp
EFFICIENT ALLOCATION OF THE INTERNET RESOURCES AND ITS PRICING

MAKOTO TAKASHIMA

OVERVIEW
With the use of the Internet being rapidly proliferated among people, congestions in transmission lines and other facilities and delays in communications have brought the topics of efficient utilization of network resources and consumers’ welfare to great concerns of economists in the field of telecommunications-related problems. This paper discusses the relation of congestion pricing with the welfare of the Internet society with the use of a model of a basic network structure characterized by hierarchical connection between backbone and service providers, and shows that the pricing scheme internalizing adverse effects of externalities on network can lead to the optimal allocation and utilization of the resources in realizing the maximal social welfare in this layered structure run by a co-operative agent and can be complied with economic behaviors of both users and providers of the Internet services in competitive circumstances as well.

1. INTRODUCTION

The Internet was initiated by ARPANET, the computer networking program of the U.S. Department of Defense in the late 1960s and developed by succession of NSFNET, the program of connecting supercomputer centers created by the NSF around the United States. Those years, the network was run by a non-profit organization under the NSF award for scientific projects and the use was restricted to non-profit activities. Thence, a new administrative non-profit entity, Advanced Network Systems, was instituted with the contribution of Merit, IBM and MCI for the resources in order to meet the demand from for-profit concerns. This development has produced the general practice of the Internet pricing that users pay connection fees only and are allowed for unlimited usage of the network for transmission°.

It was easily anticipated that we were to face the classical ”problem of the commons” in that the overgrazing of the Internet resources would produce their inefficient utilization and allocation and deterioration of users’ utility due to congestion as the use was continued

This work was supported by the Telecommunications Diffusion Foundation under Fiscal Heisei 8 (1996) Research Grant.

in this way. In fact, the NSFNET backbone experienced significant congestion in 1987, only two years after when it launched its program, and after that, as the Internet came into wide use with the free entry of providers, it happened that a local network as a whole suffered temporary suspension of working. Nowadays, it is rather usual that we experience congestion of communications and traffic delays in some parts of the Internet.

This problem is caused, at least economically, by the above described pricing and cost bearing practice shaped in the development process of the Internet which had been nurtured mainly by public subsidization and private contributions for the nonprofit purposes. Moreover, its functional or technical features make the solution difficult with inherent relations to the costbearing issues. The Internet involves technological revolutions which can be called a change of paradigm in telecommunications network. Brian Kahin gives the Internet a functional definition that “it encompasses the set of interconnected and interoperating networks which make use of the internet protocol (IP) and a common addressing scheme” (2). Accordingly, it is generally understood that the Internet is a network of networks, but in order to recognize the new paradigm, attention should be given to its specific method of data transmission, i.e., “packet exchange”, which is essentially different from that of traditional communication networks such as telephony. This new technology has realized efficient utilization of scarce telecommunication resources and enhancement of users’ benefits by increasing network externalities owing to its nature of openness which enables the network to expand freely. The new transmission technology using packet exchange makes networks “connectionless”, enabling users to share the scarce bandwidth more efficiently by allocating it randomly to the packets of any data and any senders in the way called “statistical multiplexing”. This means that the Internet treats all the packets from any data, e.g., e-mail, real-time message, urgent text, or audio-visual contents, in the equal fashion, no matter what content the data may have and what urgency it may require.

In the circumstances, while the Internet is regarded as a very promising innovative technology expected to contribute to the human society in the coming century, there have increasingly appeared the social costs of externalities caused by overuse of the network resources, and thus a field of research called “Internet Economics” is budding out these years with main interests in the pricing schemes and efficient resource allocation and usage in the Internet markets. A series of recent works by Jeffrey MacKie-Mason and Hal Varian deserves special attention among others of this field. They evaluate social costs caused by the increase in the traffic volume in congestible networks with connection to the technological features of the Internet, and analyze appropriate pricing mechanisms for

---

realization of efficient allocation of the network resources. As an operable devise in the real
transactions, "they propose a scheme called a smart market": the market is cleared
periodically by the cut-off price which internalizes the congestion costs and is set to be
compared with users' bids for packet transmission (3). Moreover, they discuss that increase
in traffic originated by an individual user gives birth to social costs of degradation of
utility of all the other users due to the rise in congestion caused on the network resources
by that traffic increment, and explain that introduction of the scheme of congestion price
reflecting the effects can encourage efficient utilization of resources and lead to optimal
social conditions in a single layered network model (4).

Following the studies made by MacKie-Mason and Varian, William Lehr and Martin Weiss
extend their modeling framework to encompass any number of networks on a single
network domain through which packets travel in various ways. According to this
generalized model, they show that congestion prices presented by the former authors proved
socially optimal in general and that the prices can be formed as addition of those for local
"on-net" traffic of each network through which the "internet" traffic travels; i.e., the
optimal congestion prices for any traffic can be set on the basis of only local information
of costs and traffic (5).

The work of Lehr and Weiss should be highly rated in that they showed using a general
"network of networks" that congestion pricing becomes generally the optimal scheme for
maximization of social welfare of the network. However, their analysis is for a set of
homogeneous networks having no price settlements with foreign networks that should
accompany the expansion beyond the boundaries: nowadays, connection with foreign
networks is an essential ingredient of the structure of the Internet. Besides, it does not deal
with sharp competition among for-profit providers in an explicite manner in their model:
such a competitive structure is also a feature of today’s Internet market. In this sense, as
Shenker et al point out, their modeling framework can be said "separate situation setting"
and considered as a model for "a single non-profit research network" or "networks within
a single cooperative organization" (6).

In this paper, we set a framework representing a basic network structure of the Internet at
present and analyze a contemporary "problem of the commons" that an individual's usage
of networks for his or her own benefit prevents Internet society as a whole from making
efficient utilization of the scarce telecommunications resources, making account of what

The analyses and results of MacKie-Mason = Varian and Lehr = Weiss can be viable under our layered structure with international connection and how they should be amended if not viable. We have already observed the rapid diffusion of the Internet in Japan these two or three years and have made a detailed inquiry into how the "network of networks" has been constructed along with the development of the market\(^{(7)}\). The whole network consists of networks supplied by many Internet service providers (ISPs), and for the analytical purpose, they may be considered to be divided into two large groups, i.e., backbone providers (BPs) and service providers (SPs). BPs are providers situated at a higher layer of the whole network which have their (own or leased) transmission circuits and supply other providers with connection service to their circuits while offering Internet services to their own end-users. Among them, there are those who have direct connections beyond country borders with foreign networks. These providers offer transmission services of backbone to the whole network, and thus they can be called backbone providers. Another group is that of providers situated at a lower layer, having connection with one or more providers of the former group. Their main business is to supply general users with connection and other Internet-related services. We name entities belonging to this group service providers in this paper\(^{(8)}\).

Succeedingly, in Section 2, we present a basic model of a network structure and behavioral presumptions about users and providers within it as the analytical framework here, based on the Internet structure in Japan (almost the same in other countries) described above. In Section 3, congestion prices are constructed so as to internalize social costs caused by network usage according to MacKie-Mason and Varian, and it is examined to what extent they can be established in the cooperative market structure under our layered network model expanded to include foreign connections. Moreover, account is taken of the conditions under which the capacity should be expanded for both SP and BP networks from the viewpoint of the total welfare of the Internet society. Section 4 analyzes the results in the industrial structure where both of SP and BP providers offer their services independently, as for-profit firms, and contrasts them with those under the cooperative organization in the previous section. While analyses up to this section are all under the assumed structure of supply side having two representative entities of a SP and a BP, it has become an actual environment of the market today that a host of private providers having various technological and business features offer services competitively to other providers and/or users for their profits purpose. Under this situation, it would be of natural need to inquire into whether and to what extent the congestion pricing should be useful in this

---

\(^{(8)}\) As a matter of fact, it would be hard to classify clearly each of the actual entities into either group. This classification is for the analytical purpose only, here.
competitive structure of our extended model. Section 5 discusses this problem. We summerize all the results obtained in this research and give in relation to them a perspective of the lines of further work to be tackled with, in Section 6.

2. INTERNET COMMUNICATION MODE

First of all, we explain the network structure of the Internet, features of the networks, types of traffic, and preferences of users.

The Internet actually consists of complex interconnections with hierarchical structure between many Internet service providers of various sizes and technical characteristics, eventually offering connection and other services to numerous end-users. We classify these providers into two groups; upper layered group of backbone providers (BPs) and lower layered group of service providers (SPs). A representative provider is sorted out from each group of BPs and SPs and two networks of these providers at different layers are connected to each other. The selected entity BP has overseas connection with foreign networks but has no end-users for itself; i.e., it offers only backbone transmission service to the other provider SP. SP supplies end-users with connection and other Internet-related services and sends data with foreign destinations to BP to have them transmitted beyond boundaries since it has no connection with foreign entities for itself; i.e., SP offers Internet services to users and receives backbone service from BP. The present capacity (maximum bandwidth) of BP is $K_B$ and that of SP is $K_s$. All these are the features of the network structure on which we proceed to consider the problem henceforth.

Next, we specify the details of end-users of the Internet services (we simply call them "users" hereafter). They are divided into two groups by types of network usage. One is a group of users sending data which both originate and terminate on the SP network domain, and the other is a group of users sending data which originates on SP network and is transferred overseas through BP network(9). We call the former users domestic communication users (DC users) and the latter ones foreign communication users (FC users), and the respective data are named DC data (or traffic) and FC data (or traffic), correspondingly. It is assumed that there are $m$ DC users and $n$ FC users, who are designated a number from $i=1$ to $i=m$ and from $j=1$ to $j=n$, respectively. Traffic volume which DC user $i$ sends is $x_i$ and that of FC user $j$ is $y_j$. Then, the total volume of DC traffic amounts to $X(=x_1+\ldots+x_m)$ and that of FC traffic becomes $Y(=y_1+\ldots+y_n)$. In addition of these

(9) Using the terminology by Lehr, W.H., and M.B.H. Weiss (1996), the former data is called "on-net" traffic and the latter corresponds to "internet" traffic. A user actually making both types of usage may be treated as two different users each of whom makes a single type of them. See note 4 of Lehr and Weiss (1996).
traffic X and Y originating on the SP network, we have traffic originating on the overseas network. This enters our network domains through the channel connected between our BP network and the foreign one, and travels on them to users having subscription with the SP network. We call this overseas communication (OC) traffic. The volume of this traffic from overseas users Z is determined by the conditions outside our network and presumed to be a given constant value.

As to the traffic on each of the network domains, three kinds of traffic of DC, FC and OC data travel past each other on the SP network domain, and two kinds of traffic going out for foreign destinations (FC data) and coming in for domestic users (OC data) are on the BP network. Hence, denoting the total volumes of traffic on the SP and the BP network domains by \( T_s \) and \( T_b \), respectively, we have \( T_s = X + Y + Z \) and \( T_b = Y + Z \). Then, the values of capacity utilization of those networks, i.e., degree of congestion of resources, \( R_s \) and \( R_b \), are described as \( R_s = T_s / K_s \) and \( R_b = T_b / K_b \).

That is the basic model of the Internet for the analysis in this paper. Subsequently, we explain the relations of payments between users and SP and between SP and BP concerning the demand and supply of the Internet services in this network. Firstly, we assume "two-part tariff" for fees which users pay to their provider SP, i.e., a constant subscription (or initial connection) charge \( S \) plus fees based on usage (packets sent). We assume that different prices are set for packets of on-net (DC) traffic and internet (FC) traffic, and denote them by \( P_s \) and \( P_y \), respectively. Hence, payments of a DC user i become \( S + X_i \cdot P_s \), and those of a FC user j are \( S + Y_j \cdot P_y \).

The transaction of SP is like this: it receives the above service fees from m DC users and n FC users, and pays the provider BP the backbone service fees of a two-part tariff consisting of the initial connection charge of \( B \) and use fees depending on the volume of traffic to be transferred; \( B + Y \cdot P_s \). Here, \( P_s \) is a price set for SP to ask BP to transfer internet traffic from its customers to the overseas destinations through backbone circuits. BP receives the service fees from SP and, besides, has exchanges of payments with foreign entities in regard to international communication under their settlement agreement.

In actuality, this settlement concerning the share of revenue from international communication is carried out by bilateral talks between the entities at each end of a specific connection route. An "accounting rate" is decided for international communication through a given route between those two carriers concerned and then the share for each party is negotiated. As usual, the agreed accounting rate is divided in half, but sometimes unequal splits are introduced on some political or economical reasons. This final agreement for
share of an accounting rate is called a "settlement rate", and each international carrier pays for traffic originating on its side and receives fees for traffic sent from the other side, based on this settled unit price. Concerning the settlement on charges of international communication, there usually exist difficult problems arising from the imbalances in the flow of traffic between two parties concerned. It may be an interesting subject to be discussed in the field of telecommunications-related studies, but it would not be suitable for us to deal with the problem here further in relation to the subject of this paper.

Considering the actual system of sharing revenue from overseas communication, we assume that the accounting rate is split in half to decide the shares of revenue between BP and its foreign partner, and a unit price for traffic for both sides is settled to be $P_f$ (settlement price). Then, BP pays its foreign partner net fees of $P_f \cdot (y - Z)$. On the other hand, BP pays SP the sum of a part of fees received $Z \cdot P_f$ in order to have the traffic entrusted by its overseas entity delivered to the final destinations, i.e., domestic users. This is a payment to be handed over to SP and assumed to be $Z \cdot P_s$. Here, $P_s$ is a unit price for the acceptance and forward delivery of overseas originating traffic by SP.

In the analytical framework described so far, we have some kinds of prices. While the subscription or initial connection charges $S$ and $B$ and the settlement rate $P_f$ are assumed given values, all the other prices are considered to vary, reflecting qualities of data transmission or networks. Unit prices of the part of use-based fees which users pay to SP are determined in relation with qualities of transmission; $Q_x$ for on-net (or DC) traffic and $Q_y$ for internet (or FC) traffic. On-net traffic originates and terminates on the same SP network and hence $Q_x$ depends on the degree of congestion of the SP network $R_s$. We assume that there is a relationship between them represented by a differentiable, decreasing convex function $Q_x = D(R_s)$. Regarding the quality of internet traffic $Q_y$, it can be considered sum of the effects of congestion of both SP and BP networks since the traffic originates on the SP network and goes through the BP domain for the foreign destination. Therefore, the relationship between the quality and the network congestion may be written as $Q_y = D(R_s) + D(R_b)$ using the same functional form $D$ as $Q_x$. Other prices, $P_s$ and $P_b$, are determined by the qualities of network resources of SP and BP, $Q_s$ and $Q_b$, respectively, and they are related with each other in the same way; $Q_s = D(R_s)$ and $Q_b = D(R_b)$.

Finally, we specify utility functions of users concerning usage of the Internet. While the utility of a user naturally increases with the volume of traffic he or she sends, it will decrease with a rise in his or her unsatisfaction due to traffic congestion or delays of communication. Thus, a user's utility for use of the Internet can be represented as a function of his or her volume of traffic and the quality of communication; i.e., $u_i(x_i, Q_x)$
for SC user $i$, and $v_j(y_j, Q_y)$ for FC user $j$, where both $u_i$ and $v_j$ are differentiable and increasing concave functions. Furthermore, it is generally considered that each user does not know how many other users share the same network resources and how they use them (e.g., volume, content, frequency, urgency, etc.) and that a user ignores the effect of congestion his or her traffic may impose on others. That is, the congestion effect caused on a user's utility by an increase in his or her traffic may be regarded as almost zero, in general.

Under the framework of the Internet society described above, we consider the problem of the pricing scheme and the efficient utilization of network resources which can lead to the enhancement of the whole welfare of the information society. In the next section, we introduce congestion pricing to internalize the social costs of usage and analyse its relations with users' behavior and the social welfare under a cooperative management of the network as the first industrial circumstance.

3. CONGESTION PRICING AND EFFICIENT RESOURCE USE

(1) Traffic Congestion Pricing

Under the general descriptions and assumptions of the analytical framework given in the preceding section, we specify here the industrial situation that we firstly deal with. The network society is constructed by two entities connected with each other, a service provider SP and a backbone provider BP, and $m$ DC users and $n$ FC users, each subscribing to SP. BP has a connection with foreign networks and offers only backbone service to SP. The whole network is run by a sort of cooperative organization consisting of those providers and users, as the historical development of the Internet.

It is generally recognized that pricing schemes without relation to usage (typically, "flat-rate" pricing) are prone to lead to the so-called "problem of the commons" by "over-grazing" the scarce resources. Actually, there always appear traffic congestion and communication delays in some parts of the network, and the adverse effects of overuse have been further intensified these days. Under the circumstances, we presume that the society thinks to resolve this problem by introducing a pricing scheme of imposing the social costs on users depending on usage of the resource, i.e., "usage-based pricing". However, it does not realize efficient use of the network resources to set a constant unit price for a packet regardless of the variable condition of network congestion: the most part of the cost of communication network resources is the fixed cost of equipments like lines, routers and other switching facilities, etc., and hence, when the network is not congested and large...
bandwidth is available, free charge for traffic would be rather preferable for enhancement of efficiency in resource use and should lead to the increase in social welfare. On this account, the society decides to introduce a usage-based pricing scheme but to make a unit price variable according to the network congestion.

Network use by a user causes congestion, to some extent at least, on the common resources of lines, routers and other communication facilities, leading to a fall in benefits of all the other users due to the delays of communication or drops of packets. The aim of setting prices reflecting the network congestion is to induce users to decide their behavior on a comparison between the value of sending information and the price for its transmission and to realize efficient use of scarce network resources of the society in compliance with economic motives of individual users in order to cure the drawbacks of fixed pricing schemes. When a user sends more information, it increases congestion on the networks through which the information travels with the consequent effects of falling the quality of communication, resulting in the decrease in utility of all the other users. Hence, it is conceived that the decrease in utility is the social cost incurred by the use of network and we may devise a pricing scheme which imposes the cost on the user originating the extra traffic. When the network is congested, a user would dare send his or her information as long as he or she considers the information has a value comparable to the high price at that time. Otherwise, the user would refrain from sending it and await the network to be uncongested. Under these considerations, prices for domestic communication (SC) traffic and foreign communication (FC) traffic are written as:

\[ p_x^* = - \sum_{k \in m, x \neq i} \left( \frac{\partial u_k}{\partial Q_x} \right) \left( \frac{\partial Q_x}{\partial x_i} \right) - \sum_{j \in n} \left( \frac{\partial v_j}{\partial Q_y} \right) \left( \frac{\partial Q_y}{\partial x_i} \right) \]

\[ = - \sum_{k \in m} \left( \frac{\partial u_k}{\partial Q_x} \right) \left( \frac{\partial Q_x}{\partial x_i} \right) - \sum_{j \in n} \left( \frac{\partial v_j}{\partial Q_y} \right) \left( \frac{\partial Q_y}{\partial x_i} \right) \]

\[ p_y^* = - \sum_{l \in n, i \neq l} \left( \frac{\partial v_l}{\partial Q_y} \right) \left( \frac{\partial Q_y}{\partial y_j} \right) - \sum_{l \in m} \left( \frac{\partial u_l}{\partial Q_y} \right) \left( \frac{\partial Q_y}{\partial y_j} \right) \]

\[ = - \sum_{l \in n} \left( \frac{\partial v_l}{\partial Q_y} \right) \left( \frac{\partial Q_y}{\partial y_j} \right) - \sum_{l \in m} \left( \frac{\partial u_l}{\partial Q_y} \right) \left( \frac{\partial Q_y}{\partial y_j} \right) \]

Although the above equations are originally constructed as the prices set for a particular SC user i and FC user j, respectively, these are the prices commonly imposed on all users,
since the effects of the extra traffic on quality of communication are not different with the sender and we have an assumption of users' ignorance about the effects of their transmission on their own utility. We proceed to analyse in the next subsection how this pricing scheme will effectuate efficient use of the network resources and how the prices can be evaluated in the real world if it deserves to be adopted.

(2) Behavior of Internet users

We first consider how individual users decide their communication volume under the prices set in the above way. Since they decide it so as to maximize their own utility from using the network, the principle of their behavior becomes as follows:

\[
\text{Max}_{x_i}[u_i(x_i, Q_x) - S - x_i \cdot p_x^*]
\]

From this, we obtain the next equation as a first-order condition:

\[
\frac{\partial [u_i(x_i, Q_x)]}{\partial x_i} - p_x^* = 0 \quad 3-(3)
\]

That is, a SC user i should decide his or her communication volume in such a way as the marginal utility by additional usage is equal to its price.

The next problem of the same kind as the above faces a FC user j:

\[
\text{Max}_{y_j}[v_j(y_j, Q_y) - S - y_j \cdot p_y^*]
\]

This optimization problem gives a first-order condition:

\[
\frac{\partial [v_j(y_j, Q_y)]}{\partial y_j} - p_y^* = 0 \quad 3-(4)
\]

which has the same meaning as the case for a SC user i.

The conditions derived above show that any user's choice is to send the volume to such an extent to satisfy the equation 3-(3) or 3-(4) according to the variable situation, no matter how the price is set. Denoting the communication volume of SC user i and that of FC user j as \(x_i^*\) and \(y_j^*\), respectively, we know that the whole volume of traffic on the network is \(X^* = x_i^* + \ldots + x_n^*\) for on-net data on the SP network domain and \(Y^* = y_1^* + \ldots + y_s^*\) for internet data on SP and/or BP network domain(s) in the user equilibrium.
From the standpoint of the network organization, in turn, is this state of network usage favourable to the society as a whole? In other words, does it realize desirable use of resources for the whole Internet society including both SP and BP providers as well as \( m+n \) users?

(3) Social Optimality of the Whole Network

We consider the optimal use of the network resources under the existing capacity \( K_b \) and \( K_s \) from the standpoint of this Internet society as a whole. The optimality here means the situation where the society attains the maximal welfare and the social welfare is measured by the total benefits net of the costs gained by all the entities concerned including users, i.e., net profits of SP and BP providers and net benefits of \( m+n \) users. This is expressed as a function \( W \) of communication volumes of each user,

\[
W(x_1, \ldots, x_m, y_1, \ldots, y_n ; K_s, K_b)
= \sum_{i \in m} u_i(x_i, Q_x) - S - x_i \cdot p_x \\
+ \sum_{j \in n} v_j(y_j, Q_y) - S - y_j \cdot p_y \\
+ \{ B + [Y \cdot p_B - Z \cdot p_s] - C_B(K_b) - p_F(Y - Z) \} \\
+ \{(m+n) S + [\sum_{i \in m} x_i \cdot p_x + \sum_{j \in n} y_j \cdot p_y] \\
- B - C_s(K_s) - [Y \cdot p_B - Z \cdot p_s] \}
\]

In this welfare function, the first and second terms of the right side of the first equation are the total value of net benefits from the network use of \( m \) DC users, \( x_i(i=1, \ldots, m) \), and that of \( n \) FC users, \( y_j(j=1, \ldots, n) \), respectively. The third term represents net profit of backbone provider BP, and the fourth is that of service provider SP. In these terms, \( C_B \) and \( C_s \) are the costs of network facilities realizing the communication capacity (maximum bandwidth) of each network, and they are represented as functions of the capacity \( K_b \) and \( K_s \), respectively, which are assumed to be differentiable and increasing.

In order to make the social welfare of the Internet society as great as possible, the organization tries to decide the optimal supply of bandwidth to each of the users. The welfare maximization problem facing the organization gives a first-order condition for a DC user as follows:
\[
\frac{\partial W}{\partial x_i} = \frac{\partial [u_i(x_i, Q_x)]}{\partial x_i} \\
+ \sum_{k \in m, k \neq i} \frac{\partial [u_k(x_k, Q_x)]}{\partial x_i} \\
+ \sum_{j \in n} \frac{\partial [v_j(y_j, Q_y)]}{\partial x_i} \\
= \frac{\partial [u_i(x_i, Q_x)]}{\partial x_i} \\
+ \sum_{k \in m, k \neq i} (\frac{\partial u_k}{\partial Q_x})(\frac{\partial Q_x}{\partial x_i}) \\
+ \sum_{j \in n} (\frac{\partial v_j}{\partial Q_y})(\frac{\partial Q_y}{\partial x_i}) \\
= 0 \quad 3-(6)
\]

Applying the congestion price formulated as the equation 3-(1) to this condition, we have the following relation:

\[
\frac{\partial [u_i(x_i, Q_x)]}{\partial x_i} - p_{x^*} = 0 \quad 3-(7)
\]

This corresponds with the condition of user’s equilibrium for the utility maximization obtained in equation 3-(3) for a DC user i. It means that the adoption of the pricing scheme internalizing the adverse effects of externalities by network congestion is not contradictory to users’ behavior of maximizing their net benefits from the network use and that it could lead to an efficient allocation of the scarce resources.

How about foreign communication through the backbone network? Differentiating the social welfare function W with respect to a FC traffic \(y_n\), we have a first-order condition for the maximization problem:

\[
\frac{\partial W}{\partial y_j} = \sum_{i \in m} \frac{\partial [u_i(x_i, Q_x)]}{\partial y_j} \\
+ \frac{\partial [v_j(y_j, Q_y)]}{\partial y_j} \\
+ \sum_{i \in n, i \neq j} \frac{\partial [v_i(y_i, Q_y)]}{\partial y_j} - p_F \\
= \sum_{i \in m} \left( \frac{\partial u_i}{\partial Q_x} \right) \left( \frac{\partial Q_x}{\partial y_j} \right) \\
+ \frac{\partial [v_j(y_j, Q_y)]}{\partial y_j} \\
+ \sum_{i \in n, i \neq j} \left( \frac{\partial v_i}{\partial Q_y} \right) \left( \frac{\partial Q_y}{\partial y_j} \right) - p_F \\
= 0 \quad 3-(8)
\]

Putting the congestion price for a FC user \(P'_F\), this equation is rewritten in the following simple form:
\[
\frac{\partial [v_j (y_j, Q_y)]}{\partial y_j} - p_y^* - p_F = 0 \quad 3-(9)
\]

This relation obtained from the social optimization problem differs by a settlement rate between provider BP and its foreign partner, \( P \), from the equilibrium condition for FC users

\[
\frac{\partial [v_j (y_j, Q_y)]}{\partial y_j} - p_y^* = 0
\]

which has already been shown in 3-(4).

It is known from the nature of users' utility functions, i.e., concavity with respect to the volume of traffic, that if the network is run under the original price \( P \) applied to FC traffic, the total volume of the traffic will exceed the optimal value to realize the efficient use of the network resources, resulting in a sort of "the problem of the commons" due to "over-grazing" by FC users. If the organization wants to maintain the optimal use of the network resources along with admitting such choices of individual users, it will need to seek subsidies of the amount of \( P_i \cdot (Y-Z) \) from outside of the Internet society. Consequently, if the organization likes to stay in a self-supporting entity, it will be necessary for them to internalize with FC price the externalities caused by the additional congestion produced by such overuse of the FC users.

The increase in FC traffic by \( d[y_j] \) imposes an additional cost of \( d[P_i (Y-Z)] \) on the network or the whole society including the network in terms of the decrease in the level of the network welfare due to the rise in congestion or the subsidy to be borne by the outside world. Hence, it will be a reasonable resolution of this problem to impose the cost on the FC users. That suggests using a revised congestion price for FC traffic, \( P_y^{**} \), instead of the original one, \( P_y^* \), :

\[
p_y^{**} = -\sum_{i \in n, i \neq j} \left( \frac{\partial v_i}{\partial Q_i} + \frac{\partial Q_i}{\partial y_j} \right) - \sum_{i \in m} \left( \frac{\partial u_i}{\partial Q_x} + \frac{\partial Q_x}{\partial y_j} \right) + \frac{d[p_F (Y-Z)]}{dy_j} \\
= p_y^* + p_F \quad 3-(10)
\]

Under this revised pricing scheme, the first-order condition for the social welfare maximization with respect to FC traffic is rewritten as

\[
\frac{\partial [v_j (y_j, Q_y)]}{\partial y_j} - p_y^{**} = 0 \quad 3-(10)
\]
which comes to correspond with the equilibrium condition for utility maximization of individual FC users. In this case, FC traffic price \( P_f' \) in the net benefit equation leading to the condition 3-(4) is presumed to have changed to \( P_f'' \) at the beginning.

The above analysis shows that the optimal use of the whole network can be effectuated in compliance with individual choices of users based on utility maximization by introducing the original congestion price, \( P_x' \), for domestic communication and the congestion price revised to accommodate the additional social cost incurred by overseas transfer of traffic, \( P_y'' \), for foreign communication. This means that, regarding Internet pricing, a principle of internalizing the adverse effects of use on the network through traffic prices will serve solution of "the problem of the commons" in a more general sense than MacKie-Mason = Varian and Lehr = Weiss have suggested in their single layered models.

Finally in this subsection, we inquire into the relation between \( P_x' \) for on-net (DC) traffic and \( P_y'' \) for internet (FC) traffic derived above.

Using the functional relations between communication quality and traffic volume explained in the descriptions of the modeling framework in the first subsection, we arrange the expressions for those prices. We recall the relations

\[
Q_x = D(R_s) \\
Q_y = D(R_s) + D(R_B) \\
R_s = (\sum_{i\in m} x_i + \sum_{j\in n} y_j + Z)/K_s \\
R_B = (\sum_{j\in n} y_j + Z)/K_B
\]

Applying these to the expression for domestic traffic 3-(3), the congestion price is rewritten as follows:

\[
p_x = -(1/K_s) \left[ \sum_{k\in m, k\neq l} (\partial u_k/\partial Q_k) (\partial Q_x/\partial R_s) + \sum_{j\in n} (\partial v_j/\partial Q_y)/(\partial Q_y/\partial R_s) \right] 3-\alpha
\]

The revised congestion price for foreign communication 3-(10) is also expressed as

\[
p_y'' = -\sum_{i\in n, 1 \neq j} (\partial v_i/\partial Q_y) \left[ (\partial Q_y/\partial R_y) (1/K_s) + (\partial Q_y/\partial R_B) (1/K_B) \right] \\
- \sum_{i\in m} (\partial u_i/\partial Q_x) (\partial Q_x/\partial R_s) (1/K_s) + p_f
\]

in a similar way.
It may be assumed that an increase in traffic an FC user sends $y_i$ will have only a negligible effect on a certain DC user's utility $u_i$ through the fall in quality of domestic communication $Q_x$ since the number of users, $m$ and $n$, are very large. That is,

$$(\partial u_i / \partial Q_x) (\partial Q_x / \partial y_i) \approx 0$$

Using our former presumption of user's ignorance about the adverse effect of own traffic along with this assumption, we have the following relation:

$$p_Y^{**} \approx p_Y^* + p_F - \sum_{i \in n} (\partial v_i / \partial Q_y) (\partial Q_y / \partial R_b) (1 / K_b)$$

Here, the third term of the right side of the equation expresses external diseconomies imposed on all the FC users through the rise in congestion on BP network caused by a marginal increase in traffic volume originating from an FC user ($VYBY$).

(4) Revised Congestion Pricing and Conditions for Capacity Expansion

With the use determined by the above institutional situation, the network realizes the efficient resource allocation and the maximum social welfare in terms of the total net benefit of the entities concerned under given capacity of each network of BP and SP. Then, the maximum value of the net benefit will vary according to the network capacities, and it can be represented by a maximum value function of variables, i.e., $W(K_b, K_s)$. Using this function, we go on to consider the problem of how the organization can know the situation to expand capacity.

We begin with the SP network. If the organization expands that capacity $K_s$, it reduces the utilization rate of the SP network. This causes in turn the rise in use of each user on account of the increasing benefits from use of the Internet. With this effects in mind, differentiating the welfare function 3-(5) with respect to $K_s$, we have the equation of the first-order condition for the optimal capacity:

$$\frac{\partial W}{\partial K_s} = \sum_{i \in m} \frac{\partial [u_i (x_i, Q_x)]}{\partial K_s} + \sum_{j \in n} \frac{\partial [v_j (y_j, Q_y)]}{\partial K_s}$$

$$- \frac{dC_s}{dK_s} - p_F \cdot \sum_{i \in n} \frac{\partial y_i}{\partial K_s} = 0$$

Envelop theorem tells that the first-order conditions 3-(6) and 3-(8) with respect to $x_i$ and $y_i$ should be still established at this equilibrium point, and hence, using those relations and
the equation of the congestion price 3-(12) for DC traffic, we can arrange the above equation as follows:

$$\frac{\partial W}{\partial K_s} = R_s p_x^* - \frac{dC_s}{dK_s} = 0$$  \hspace{1cm} 3-(40)

As to the capacity of the BP network, using the equation of the revised congestion price 3-(13) for FC traffic as well as the same conditions 3-(6) and 3-(8) as the above, the similar arrangements give the following equation:

$$\frac{\partial W}{\partial K_B} = R_B \left[ - \sum_{i \in n} \left( \frac{\partial v_i}{\partial Q_Y} \right) \left( \frac{\partial Q_Y}{\partial R_B} \right) \right] \left( \frac{\partial R_s}{\partial y_i} \right) - \frac{dC_B}{dK_B} = \left( p_y^{**} - p_x^* - p_F \right) - \frac{dC_B}{dK_B} = 0$$  \hspace{1cm} 3-(45)

The above results tell us the situation where the network capacity must be expanded. As regards the SP network domain, the capacity expansion will increase the total net benefit when the cost of total traffic on the SP domain measured by congestion price for DC traffic, \((X + Y + Z) \ P_s^*\), comes to exceed the value of the capacity measured by the marginal cost of capacity, \(K_s(dC_s/dK_s)\). As regards the BP network domain, increasing capacity will become meaningful when the total social cost incurred on the BP domain by the marginal increase in FC traffic exceeds the value of capacity measured by the marginal cost of capacity, \(K_b(dC_b/dK_b)\). That total social cost can be measured alternatively by the cost of total traffic on the BP domain valued by the revised congestion price for FC traffic net of the congestion price for DC traffic and the settlement rate with the foreign network, \((Y + Z)(P_y^{**} - P_x^* - P_F)\).

Turning back to the expressions 3-(12) and 3-(13) for our revised congestion prices, it would be difficult to measure them on the basis of utility functions of individual users and quality functions of communications since in general they are not known in the explicit form. With the above relations for the optimal capacity being established, however, the prices can be rather easily found only by the information about the volume of traffic presently existing on each of the network domains and the values of their capacity as expressed in the following equations:

$$p_x^* = \left[ K_s \left( \frac{dC_s}{dK_s} \right) \right] / (X + Y + Z)$$  \hspace{1cm} 3-(46)

$$p_y^{**} = p_x^* + p_F + \left[ K_b \left( \frac{dC_b}{dK_b} \right) \right] / (Y + Z)$$  \hspace{1cm} 3-(47)
That is, DC price can be calculated on the information about only SP network, and FC price can be obtained by adding the BP capacity value per traffic on that domain and the settlement rate to the DC price.

4. INDUSTRIAL ORGANIZATION(1) : MONOPOLISTIC SUPPLY

So far, we have analyzed the use of the Internet in the industrial situation where the network is managed by a sort of cooperative organization after the historical fact of the development. As already described, the use has become wide-spread among the general public with for-profit providers entering the industry as suppliers of various Internet-related services as well as network connection. Reflecting this reality, we are going to analyze the relation of congestion pricing to choices of the private entities concerned and examine to what extent the desirable results obtained in the non-profit environment will be maintained in the industrial organizations consisting of for-profit providers in the present section and the next. We deal with the industrial organization where services are supplied monopolistically in this section and analyse the case where they are supplied competitively by many private providers in each layer of SP and BP networks in the next section. In either case, we assume that there always exist numerous users with various tastes and preferences with respect to Internet use.

The supply side in this section consists of an SP firm supplying connection service to users and a BP firm offering backbone service to the SP, and they make choices for maximization of their own profits. The other things remain unchanged. We assume that all the FC traffic originating from SP domain is accepted and delivered to the foreign network by BP provider.

In this monopolistic situation, users have no choice of their own providers and decide their use, i.e., volume of traffic, in consideration of communication quality, Qx or Qy, presented by SP and traffic price, Px or Py, set in the market at that time, according to the type of traffic. That is, users' equilibrium conditions are given in the same behavior of maximizing net benefits as considered in the second subsection of Section 2:

$$\frac{\partial u_i}{\partial x_i} = p_x$$  \hspace{1cm} 4-1
$$\frac{\partial v_i}{\partial y_i} = p_y$$  \hspace{1cm} 4-2

Each user determines his or her traffic volume, x, or y, according to this condition, and subscribes to SP service when the net benefit becomes positive with the use of the network. However, all the traffic decided by the subscribers is not always accepted by SP provider,
because this monopolistic supplier SP can decide the total volume of supply of service on its own criterion, i.e., profit maximization. Conversely, the volume which SP decides to accept is always met by demands for service by users, since many users are presumed to always exist with a diversity of utility functions. We assume that the numbers of users who actually receive service from SP are \( m \) for DC users and \( n \) for FC users.

Next, the behavior of provider SP is to determine the volumes of supply of service, \( X \) for DC traffic and \( Y \) for FC traffic, and its transmission capacity \( K_s \) so as to maximize its net profit:

\[
\Pi_s = [X \cdot p_x(Q_x) + Y \cdot p_y(Q_y) + (m+n) S] \\
- [Y \cdot p_b(Q_b) + Z \cdot p_s(Q_s) + B + C_s(K_s)]
\]  

Although a monopoly SP could decide prices for DC and FC traffic on its own will, they are actually established in the market through the aggregate demand function with relation to the quality of communication for each type of traffic. Hence, the prices \( P_x \) and \( P_y \) are expressed as a function of the corresponding quality of communication \( Q_x \) and \( Q_y \). SP sends the FC traffic originating from its subscribers to BP in order to have the traffic delivered to the foreign destinations through BP network. \( P_b \) is a price per packet which SP pays to BP for it. \( P_s \) is a price which BP hands over to SP for acceptance and onward delivery of the international traffic originating from BP’s foreign partner. We consider that these prices are expressed as a function of the quality of networks determined by the respective degrees of congestion. Notice here that the quality of SP network, \( Q_x \), is equal to the quality of DC communication, \( Q_s \), which was already explained in connection with the description of the modeling framework in Section 2.

Under those functional relations, the net profit maximization problem facing provider SP gives the following first-order conditions:

\[
p_x = -X \left( \frac{\partial p_x}{\partial Q_x}\right) \left( \frac{\partial Q_x}{\partial X}\right) \\
-Y \left( \frac{\partial p_y}{\partial Q_y}\right) \left( \frac{\partial Q_y}{\partial X}\right) \\
-Z \left( \frac{\partial p_s}{\partial Q_s}\right) \left( \frac{\partial Q_s}{\partial X}\right)
\]

\[
p_y = -X \left( \frac{\partial p_x}{\partial Q_x}\right) \left( \frac{\partial Q_x}{\partial Y}\right) \\
-Y \left( \frac{\partial p_y}{\partial Q_y}\right) \left( \frac{\partial Q_y}{\partial Y}\right) \\
-Z \left( \frac{\partial p_s}{\partial Q_s}\right) \left( \frac{\partial Q_s}{\partial Y}\right) \\
- \partial \left[ Y \cdot p_b(Q_b) \right] \left( \frac{\partial Y}{\partial Y}\right)
\]
An increase in traffic by a DC user causes a fall in the quality of communication due to a rise in congestion of SP network, and it has an effect on all the kinds of traffic, X, Y, and Z, traveling on that domain. On that account, SP suffers a decrease in revenue due to a fall in prices. Equation 4-(4) means that the price for DC users equals the value compensating the fall in revenue resulting from an additional increase of packet in DC traffic in the situation where the monopoly SP determines the volumes of supply, X and Y (Z is given) so as to maximize its profit. On the other hand, equation 4-(5) shows that the price for FC traffic is constituted by the sum of the similar costs imposed by the increase in FC traffic and the total change in payment to BP through a fall in quality of BP network owing to the additional traffic (the second factor is the sum of the second and third terms of the right side of the equation and is designated PYBY).

Attention should be paid here to the fact that an increase in traffic by a user has the same effect on the degree of congestion of SP network regardless of the type of the traffic and hence the difference in the types makes no different effects on the quality of DC communication $Q_x$ and moreover that the quality of FC communication $Q_y$ is expressed as the sum of the qualities of both SP and BP networks, $Q_s$ and $Q_b$. Then, the price for FC traffic is written in the form, encompassing the price for DC traffic:

$$ p_Y(Q_Y) = p_x(Q_x) + \partial [Y \cdot p_B(Q_B)]/\partial Y $$

$$ -Y (\partial p_Y/\partial Q_Y) (\partial Q_Y/\partial R_B) (\partial R_B/\partial Y) $$

The second term is the value denoted PYBY above. An increase in FC traffic rises congestion of BP network and it results in a reduction in revenue of SP on account of the fall in the price for FC traffic caused by the fall in quality of FC communication. We denote the value of the fall in revenue by $YYBY$, which proves equivalent to $YVBY^{(10)}$. This is the third term of the above equation.

The task of backbone provider BP is to deliver international traffic to domestic users and to the overseas destinations through that network, and hence, although it is a monopoly supplying the backbone service, the only thing that BP decides for its for-profit purpose is its transmission capacity $K_b$. This capacity for the traffic volume traveling on the BP network $T_b$ determines the quality of that network and has influence on BP's profit through the effects on the price and the volume of FC traffic. We deal with the choice of $K_b$ by BP later, relating to the problem of SP's decision on its capacity $K_s$.

---

(10) Refer to the description leading to equation 4-38.
The whole network structure in the present analysis remains the same as what we had for the analysis up to the previous sections. Only the difference between them lies in the ways of management of the network: while the network was run under a sort of cooperative organization with the aim of maximizing the total benefit as a whole in the previous analysis, it is assumed that two mutually connected networks in a layered structure are managed by respective for-profit firms monopolistically in the section. In the present industrial organization, it is easily conceived that the profit maximization behavior of SP results in the transfer of part of users' consumer surplus to SP's profit, and above all, the total social welfare of the whole network will generally become smaller than in the case of cooperative management because of the profit maximization activities of two individual monopolistic firms in the present case.

How is the prices of 4-(4) and 4-(5) (or 4-(6)), \( P_x(Q_x) \) and \( P_y(Q_y) \), related to the revised congestion prices given in the previous cooperative model, \( P^*_x \) and \( P^*_y \)? The prices established in the monopolistic model can be considered to be higher, ceteris paribus, than those set in the the social welfare maximization of the cooperative model. Regarding the price for DC traffic, when there is an increase in DC traffic, the monopolistic firm SP will try to have the price established in the market in such a way as to compensate the reduction in revenue resulting from a rise in congestion and additionally to cover the decrease in receipts from BP concerning onward delivery of international traffic which is incurred by the fall in quality of SP network, \( Q_s \) (we denote the second cost by \( Z_{SX} \)). On the contrary, in the case where the network is managed to maximize the total benefit of the whole society encompassing both SP and BP networks, payments and receipts between both providers and between SP and users are all offset by each other in the whole society and hence they do not appear as factors having any influence on the network management.

Attention must be paid here to a difference in construction of prices between these two models: the prices in the model of monopolistic supply are found to be formed so as to compensate the fall in revenue which providers suffer owing to the rise in network congestion, while the prices in the cooperative model are set to internalize the external adverse effects which users suffer on that account. In the long-run equilibrium where users make a choice in consideration of the eventual fall in communication quality due to network congestion, the reduction in revenue of SP caused by a fall in communication quality is equal to the value of external diseconomies given to all the other users of the same communication type when a user increases use of the network. It is shown like this. As regards the choice of a DC user, the first-order condition for the net benefit maximization is derived in the similar form to the equation 3-(3) in the previous section:
\[ \frac{\partial}{\partial x_i} (x_i, Q_x) = \frac{\partial}{\partial x_i} p_x (Q_x) = 0 \quad 4-(1) \]

Here, the price \( p_x \) is not set according to the social cost of externality caused by network congestion by the cooperative organization but determined in the functional relationship with communication quality in the market. Now, when a user makes a short-run choice, taking the present communication quality to be unchanged, the above equation is written as 4-(1). However, in the long-run situation where the communication quality changes with the degree of congestion in accordance with the changes in traffic volume, the above condition derived from the optimal behavior of a user is reduced to the equation:

\[
\frac{\partial u_i}{\partial x_i} + \left( \frac{\partial u_i}{\partial Q_x} \right) \left( \frac{\partial Q_x}{\partial x_i} \right) - p_x (Q_x) - x_i \left( \frac{\partial p_x}{\partial Q_x} \right) \left( \frac{\partial Q_x}{\partial x_i} \right) = 0 \quad 4-(2)
\]

On top of that, the equation 4-(1) should be established under \( Q_x \) of the long-run equilibrium, and thus, we have

\[
\left( \frac{\partial u_i}{\partial Q_x} \right) \left( \frac{\partial Q_x}{\partial x_i} \right) = x_i \left( \frac{\partial p_x}{\partial Q_x} \right) \left( \frac{\partial Q_x}{\partial x_i} \right) \quad 4-(3)
\]

In our analysis of the effects of network congestion caused by the increase in traffic, the natural assumption is that a partial derivative of quality of DC communication with respect to DC traffic is not zero. Thus, we have (changing \( i \) into \( k \) for the convenience of later developments):

\[
\left( \frac{\partial u_k}{\partial Q_x} \right) = x_k \left( \frac{\partial p_x}{\partial Q_x} \right) \quad 4-(4)
\]

Adding this equation up across users of DC communication, \( k=1,...,m \), we obtain an expression:

\[
\sum_{k=1}^{m} \left( \frac{\partial u_k}{\partial Q_x} \right) = X \left( \frac{\partial p_x}{\partial Q_x} \right) \quad 4-(5)
\]

The congestion effect of the increase in traffic of a DC user \( i, x_i \), on quality of DC communication does not differ from that of the increase in total traffic of the same type, \( X \). Hence, a relation:

\[
- \sum_{k=1}^{m} \left( \frac{\partial u_k}{\partial Q_x} \right) \left( \frac{\partial Q_x}{\partial x_i} \right)
= -X \left( \frac{\partial p_x}{\partial Q_x} \right) \left( \frac{\partial Q_x}{\partial X} \right) \quad 4-(6)
\]

is eventually established, ascertaining what has been postulated.
For FC communication originating from SP domain, it is proved that we have the similar relation

\[- \sum_{i \in n} \left( \frac{\partial v_i}{\partial Q_Y} \right) \left( \frac{\partial Q_Y}{\partial y_i} \right) = -Y \left( \frac{\partial p_Y}{\partial Q_Y} \right) \left( \frac{\partial Q_Y}{\partial Y} \right) \]

In our original model, provider BP has a delivery contract with provider SP in a layered network structure to have overseas traffic accepted and brought to domestic users, and BP pays SP a variable price \( P_s(Q_s) \) per packet for that onward delivery. On the other hand, BP has a contract with foreign carriers and they pay a constant (for a certain period of contract) settlement rate per packet to each other. Consequently, BP's revenue on that delivery of international traffic may change with the congestion situation on SP network which is not concerned with the conditions of BP's network. Hence, it may be reasonable to make a contract of setting the \( P_s \) constant subject to the period of contract between BP and its foreign partners. When our model is partly changed in this way, \( P_s \) becomes constant with respect to quality of SP network \( Q_s \). If we further assume that BP decides the volume of supply of its backbone service (offer of bandwidth to SP) to maximize the profit, using the net profit function of BP

\[
\Pi_B = [Y \cdot p_B(Q_B) + Z \cdot p_s(Q_s)] - [p_F(Y - Z)] - C_B(K_B) + B
\]

we obtain the following relation:

\[
\frac{\partial [Y \cdot p_B(Q_B)]}{\partial Y} = p_F
\]

Applying those relations above obtained to communication prices \( P_x(Q_x) \) and \( P_y(Q_y) \) given by respective equations 4-(4) and 4-(6), we have the result that they are reduced to the revised congestion prices \( P_x^{**} \) and \( P_y^{**} \), respectively, which are set as 3-(1) and 3-(13) in the cooperative model. That is, when the delivery rate paid by BP to SP, \( P_s \), is set constant with relation to a constant settlement rate, the traffic prices established in the market of monopolistic supply become equal to the congestion prices leading to the optimal use of resources, in the situation where the optimal volumes of FC traffic decided independently by SP and BP happen to correspond with each other.

The congestion pricing was conceived to realize efficient use of network resources in compliance with the economic motives of users in a cooperative setting of the industry. From the analyses we have made so far, it's general applicability is suggested as an efficient pricing scheme in the decentralized situation where providers make a choice for
their for-profit purposes as well as in the centralized (cooperative) setting.

Finally in this section, we consider how each of providers SP and BP adjusts their network capacity to the market situation in the profit maximization behavior. For provider SP, from its net profit function of 4-(3) and the price equation 4-(4) for DC traffic, the first-order condition with respect to its capacity is derived as:

$$ \frac{\partial \Pi_s}{\partial K_s} = X \left( \frac{\partial p_x}{\partial Q_x} \right) \left( \frac{\partial Q_x}{\partial K_s} \right) + Y \left( \frac{\partial p_y}{\partial Q_y} \right) \left( \frac{\partial Q_y}{\partial K_s} \right) + Z \left( \frac{\partial p_s}{\partial Q_s} \right) \left( \frac{\partial Q_s}{\partial K_s} \right) - \frac{dC_s}{dK_s} = R_s \cdot p_x (Q_x) - \frac{dC_s}{dK_s} \tag{4-35} $$

This means that expanding capacity increases SP's net profit when the value of total traffic on the SP network measured by DC traffic price exceeds the value of its present capacity evaluated by the marginal cost of capacity. We notice that this condition is substantially equivalent to the relation 3-(14) obtained in the cooperative model.

Following a similar line for backbone provider BP, we have a relation:

$$ \frac{\partial \Pi_B}{\partial K_B} = Y \left( \frac{\partial p_B}{\partial Q_B} \right) \left( \frac{\partial Q_B}{\partial K_B} \right) - Z \left( \frac{\partial p_s}{\partial Q_s} \right) \left( \frac{\partial Q_s}{\partial K_B} \right) - \frac{dC_B}{dK_B} = R_B \left[ -Y \left( \frac{\partial p_B}{\partial Q_B} \right) \left( \frac{\partial Q_B}{\partial Y} \right) \right] - \frac{dC_B}{dK_B} \tag{4-36} $$

Here, the second term of the first equation disappears, since expansion of BP capacity does not influence quality of SP network. Thus, the equation tells that provider BP will expand its capacity when an increase in revenue produced by the improved network exceeds the cost incurred for the improvement. An increase in FC traffic rises congestion on BP network, which results in a fall in delivery rate \( p_B \) paid by SP to BP. In the second equation, the second factor of the first term means a decrease in revenue of BP effectuated by that fall in \( p_B \) rate when FC traffic increases. We denote this effect by \( YBY \). Thus, the signal for BP to expand its capacity is alternatively expressed by the situation that the value of total traffic on BP network evaluated by \( YBY \) comes to exceed the value of the present capacity.
valued using the marginal cost of capacity. It shows a difference from the situation in the cooperative model, where the value of total traffic on BP network was evaluated by the social cost of a fall in quality of FC communication (VYBY).

5. INDUSTRIAL ORGANIZATION(2) : COMPETITIVE SUPPLY

We lastly analyse a contemporary situation that many providers supply services competitively in each layer of the networks of Internet connection service (offered by SPs) and backbone service (offered by BPs), and see to what extent the results obtained so far remain viable in the new environment. When many SPs supply competitively their services in their layer of the network, qualities of communication, Qx and Qy, which they offer to users are regarded as their respective technical elements relating to their transmission capacity Ks: it is considered to be an entrepreneurial decision whether they seek to realize high prices by offering service of high quality or to have a large volume of traffic by offering that of moderate quality. On the users’ side, too, communication qualities, Qx and Qy, give them important clues to choose a provider for the maximization of their benefit in the use of the Internet. Moreover, for the group of providers BP, it becomes a basic business choice how much bandwidth Kb should be offered to providers SP for onward delivery of FC traffic originating from them, and the quality of the BP network Qb determined with relation to the offered capacity is a characteristic of each provider. Thus, a specific quality of a BP network becomes a strategic factor for SPs seeking backbone service to find their satisfactory BP. Here, we assume that there are many providers having a great variety of these qualities, i.e., with continuously different values in Qx, Qy, and Qb.

We take notice of market behavior of a representative provider belonging to each group of SPs and BPs in this circumstance. It is presumed that the selected SP provider "r" has subscribers of m DC users and n FC users, and that the selected BP provider offers backbone connection service to s firms of SP group.

Firstly, the purpose of choice behavior of SP provider "r" is to maximize the value of its net profit function 4-(3). The price equations established in the market under this behavior does not differ from those obtained in the monopolistic situation as 4-(4) and 4-(5) (or 4-(6)), and such is the case with capacity expansion. As to the optimal choice of a BP provider, the first-order condition for SP provider "r" gives the equation:

\[
\frac{\partial [Y_r \cdot p_r (Q_r)]}{\partial Q_B} = \frac{\partial [B (Q_B) + Y_r \cdot p_B (Q_B)]}{\partial Q_B} \quad 5-1
\]
Here, \( B (Q_b) \) is the initial connection fee SP pays BP, and the functional form means that it is determined with relation to the quality of BP network in the market. When SP makes a choice of a BP provider offering the network of higher quality, it will enjoy greater revenue by offering users FC communication of higher quality, while it must pay BP more for the FC delivery. The above equation means that the SP "r" will find a BP offering such network of quality as the increase in revenue brought by higher quality compensates the rise in payment to BP.

Next, we see differences in the choice of BP from that in the other industrial situations. As regards the net profit function, there is nothing different except that the constant initial connection fee \( B \) is changed into a term \( sB(Q_b) \) expressed by a number of connecting SPs and a function of quality of BP network. In the cooperative model and in a case of the monopolistic model, all FC traffic originating from SP network was accepted and delivered to foreign networks by BP. In the present situation of market, many providers having various technical characteristics offer services competitively in each layer of the networks, SP and BP. Then, for the part of BP, the amount of bandwidth to offer, i.e., the volume \( Y \) of FC traffic to accept, becomes an important choice as well as the network quality \( Q_b \) which is determined by another choice, transmission capacity \( K_b \). On the other hand, the FC traffic price itself is determined in the market in relation to demand and supply of FC communication influenced by the quality of BP network.

The profit maximization problem facing BP gives the first-order condition with respect to the volume of FC traffic \( Y \) to accept, from which we have the following equation:

\[
p_B (Q_b) = p_r + Z \left( \frac{\partial p_B}{\partial Q_b} \right) \left( \frac{\partial Q_b}{\partial Y} \right) - \left[ Y \left( \frac{\partial p_B}{\partial Q_b} \right) + s \left( \frac{\partial B}{\partial Q_b} \right) \right] \left( \frac{\partial Q_b}{\partial Y} \right)
\]

Here, \( Y \) is the sum of \( Y_r (r=1,\ldots,s) \), traffic volume of each of \( s \) firms of providers SP. Besides, as to provider "r", the relation 5-(1) should be satisfied. Thus, using these relations, we have

\[
Y \left( \frac{\partial p_Y}{\partial Q_Y} \right) \left( \frac{\partial Q_Y}{\partial Q_b} \right) - s \left( \frac{\partial B}{\partial Q_b} \right) = 0
\]

Applying this to the above equation, we can arrange the price equation as

\[
p_B^C (Q_b) = p_r - Y \left( \frac{\partial p_Y}{\partial Q_Y} \right) \left( \frac{\partial Q_Y}{\partial Q_b} \right) \left( \frac{\partial Q_b}{\partial Y} \right) + Z \left( \frac{\partial p_s}{\partial Q_s} \right) \left( \frac{\partial Q_s}{\partial Y} \right)
\]

5-2
Backbone delivery price $P_b$ for FC traffic established in this competitive situation is distinguished here by an attached letter c. In the former situation where backbone service is supplied by a monopoly BP, the delivery price was given as the equation 4-(14) when another delivery price $P_s$ for overseas traffic $Z$ is assumed constant, but when the constancy on $P_s$ is not assumed, $P_b$ is expressed as the following equation:

$$
P_b (Q_b) = P_f - Y \left( \frac{\partial P_b}{\partial Q_b} \right) \left( \frac{\partial Q_b}{\partial Y} \right) + Z \left( \frac{\partial P_s}{\partial Q_s} \right) \left( \frac{\partial Q_s}{\partial Y} \right)
$$

$P_b$ in the monopolistic situation has an ingredient of a decrease in revenue valued using the effect of BP congestion on the delivery price of FC traffic in addition to the settlement rate and the reduction in payment to SP with respect to onward delivery of overseas traffic. For the price in the competitive situation, the corresponding ingredient is valued using the effect of BP congestion on FC communication price (the second term of the equation 5-(2) is YYBY). That is, while delivery price $P_b$ is expressed by only influences on BP network exerted by an increase in FC traffic in the monopolistic situation, the influences on both BP and SP networks are included in that price when providers are in a competitive state.

The similar difference appears in the expression of the optimal capacity of BP network. From the net profit function replacing a constant initial connection fee $B$ with a function of the network quality $B(Q_b)$, the first-order condition for the maximization gives the equation:

$$
Y \left( \frac{\partial p_f}{\partial Q_f} \right) \left( \frac{\partial Q_f}{\partial Q_b} \right) \left( \frac{\partial Q_b}{\partial K_B} \right) = \frac{dC}{dK_B} \quad 5-3
$$

In the monopolistic case, we have the following relation from the equation 4-(17) in the previous section:

$$
\frac{\partial II_B}{\partial K_B} = Y \left( \frac{\partial p_f}{\partial Q_f} \right) \left( \frac{\partial Q_f}{\partial K_B} \right) - \frac{dC}{dK_B}
$$

That is, in the same way as in the case of a delivery price $P_b$ for FC traffic, in the monopolistic supply, BP's profitability of its capacity expansion is known only based on the conditions of BP network that the increase in revenue brought by an improvement of the network quality owing to the capacity expansion exceeds the cost for it. However, in the competitive situation, the economic signal for BP's capacity expansion is that BP's contribution to the increase in SP's revenue relating to FC traffic through the improvement of BP network quality exceeds the cost for the additional capacity. Here, it is natural that the capacity expansion of BP increases BP's revenue itself through a rise in delivery price.
Finally, in this competitive model, when the delivery price $P_s$ for overseas traffic is assumed fixed irrespective of SP network quality in relation to a constant settlement price $P_f$, all the results here are reduced to the same as those in the cooperative model except a price $P_b$ for FC traffic which SP pays to BP for onward delivery to foreign networks. That is, in the industrial organization where many SP and BP for-profit firms supply Internet services competitively, the communication prices established in the market essentially correspond to the congestion prices adopted in the cooperative structure to realize efficient allocation of network resources by internalizing adverse external effects of usage on the quality of communication and networks. It means that those prices are established in compliance with individual behaviors of users and providers and that they leads to efficient use of scarce resources of the whole network. This result was shown by MacKie-Mason and Varian (1955) for a set of homogeneous networks in a single layer. Our analysis has shown that the same fact is essentially established in a layered structure having providers connected with each other at different levels of networks.

6. SUMMARY AND CONCLUSION

In the Internet society, there recently have emerged external diseconomies due to line congestion and traffic delays, the extent of which is not negligible. The purpose of this paper is to analyze how the congestion pricing to internalize the external effects is viable for a layered network structure typically representing the actual market of the Internet where for-profit providers supply services competitively.

As the analytical framework, we propose a model of the Internet society consisting of numerous users with various utility functions and two kinds of providers offering Internet services, i.e., service providers BPs offering connection and other related services to end-users and backbone providers BPs supplying SPs backbone connection service only. They are connected with each other in a layered structure and BPs have channels to communicate with foreign networks. Varieties of traffic are communication originating and terminating on only SP network domains (DC traffic or communication), communication originating on a SP network and delivered to overseas networks (FC traffic or communication), and traffic originating from foreign networks and delivered to domestic users through BP networks. Delivery prices between SP and BP, $P_s$ and $P_b$, are determined by the conditions in the market, but a certain settlement rate for mutual transmission is agreed between BPs and its foreign partners.
Firstly, in the cooperative organization where the Internet is run for the welfare of the whole society, the optimal pricing for DC traffic $P_x$ is to internalize the external diseconomies caused by network congestion due to additional use of network for that traffic. As regards the pricing for FC traffic, the optimal price $P_y$ becomes the sum of the value of the similar congestion costs incurred by additional use for FC traffic and a settlement rate at that time, and has a relation with the optimal price for SC traffic expressed as $P_y = P_x + P_f + V_{YBY}$, where $V_{YBY}$ designates total costs imposed on all the FC users by a rise in congestion on BP network due to additional FC traffic. Concerning the conditions of capacity expansion, the right signal for SP network is that the value of all traffic on SP network evaluated by DC congestion price at that time exceeds the value of SP capacity measured using its marginal cost of capacity. The expansion increases BP's profit when the value of traffic on its network evaluated by the costs of congestion $V_{YBY}$ exceeds the value BP capacity evaluated by the marginal cost of capacity. From the conditions of the capacity expansion, the DC traffic price $P_x$ can be expressed as the value of SP network capacity per traffic and the FC traffic price $P_y$ can be calculated by adding $P_x$, $P_f$ and the value of BP network capacity per traffic on its network. These relations enable us to evaluate the optimal prices for traffic only on more accessible information of traffic volumes and the values of network capacity, not knowing utility functions of users and quality functions of networks.

Next, we obtain the traffic prices established in the industrial structure where each providers of SP and BP are for-profit firms independently run for respective profit maximization purposes. The network constitution itself has no change from that in the previous cooperative situation. The only difference lies in the form of management. The DC traffic price $P_x$ is obtained as the value of adding $Z_{SX}$ (reduction in delivery revenue of SP caused by a fall in SP network quality due to additional DC traffic) to the optimal DC price in the cooperative model. The FC traffic price becomes in this situation the sum of the optimal price $P_y$ in the previous situation, the additional cost $Z_{SX}$ from a fall in the network quality, and the value $Y_{BFY}$ (a change in net revenue of BP regarding FC traffic delivery due to additional FC traffic). It is equivalent to the value obtained by addition of $V_{YBY}$ and $Y_{PBY}$ (a change in payments to BP regarding FC traffic delivery brought by additional FC traffic). Concerning capacity expansion, the economic signal is given by the same relation as in the previous model, by substituting the price established in this monopolistic model for that in the cooperative model. On the other hand, as to BP network, the value of all traffic on the network is measured by the cost $Y_{BY}$ (reduction in delivery revenue of FC traffic caused by the effect of additional FC traffic on delivery price $P_b$) while measured by the cost $Y_{YBY}$ (reduction in delivery revenue of FC traffic caused by the effect of additional FC traffic on FC traffic price $P_y$) in the cooperative situation. Then,
expansion of capacity becomes profitable for BP when that value of the traffic exceeds the value of capacity. It is suggested under the general presumptions about price functions relating to congestion and cost functions of capacity that the traffic prices become higher and the optimal capacity become smaller in the monopolistic situation than those in the cooperative situation.

When we introduce the assumption that delivery price of overseas traffic $P_s$ is set a constant value between SP and BP in relation to a constant settlement rate $P_f$ agreed between BP and its foreign partner, the expression for DC traffic price is reduced to the same as the congestion price, and the condition for capacity expansion of SP network becomes equal to that in the cooperative model. As to the FC traffic price, the equation has an expression of depriving the effect caused by a change in $P_s$ of the original equation. The introduction of the assumption makes no change in capacity expansion of BP network. When we further introduce an assumption that BP decides bandwidth to supply for FC traffic, pricing expression is reduced to the same as that in the cooperative model for FC traffic, too. In the monopolistic model, a delivery price for FC traffic $P_b$ is established as the sum of a settlement rate $P_f$ and a cost $YBY$ (reduction in revenue of BP caused by a rise in BP network congestion due to additional FC traffic) in the equilibrium, and with this connection, expanding BP capacity is taken place with the signals that the value of all traffic on the network evaluated by $YBY(=P_b-P_f)$ reaches the present value of capacity.

It is generally observed that the equilibrium state in the monopolistic structure of supply is considerably different from the optimal state in the cooperative structure, but that there is little difference in results about SP network and DC communication terminating on SP domain between these two situations while the differences become greater for results about BP network and FC communication, both having a direct relation to foreign networks. However, as we introduce assumptions of a constant delivery price BP pays to SP and of a BP's decision of bandwidth to offer SP, the equilibrium state of the monopolistic model approaches the optimal situation in the cooperative model. The cooperative model is a prototype of market structure which is reached by introducing further the possibility of individual choices of all the entities concerned with the network including free entry into the market. The supply side consists of many for-profit providers with a variety of technical features of network, i.e., network qualities and communication qualities, in each group of SPs and BPs, and they offer connection or backbone service competitively to their customers, i.e., users or SPs. Thus, users' choice begins with what qualities of network and communication their provider SP should have, and SPs' first choice is also what quality of network their BP should have. On the BPs' side, in addition to the physical transmission capacity (maximum bandwidth), bandwidth to actually offer (or total traffic to accept)
becomes an important business strategy in relation to a delivery price reflecting the quality of network. In this competitive situation, the results of the economic choice by individual users and providers are found all equivalent to those obtained in the cooperative model except a delivery price of PC traffic SP pays to BP, Pb. It means that the congestion pricing introduced to improve efficiency in network resource use is basically viable in the present decentralized and layered structure on the Internet.

In the field of studies on information networks, even when focused on the relations of pricing schemes with efficient use of the resources, researches have just started with a lot of work to be done. It is difficult to conceive a line of future researches in the light of the rapid and sustained growth of information technology and the industrial development accompanying it. In the circumstances, we indicate here some of the research subjects to be tackled with and some possible lines of research.

Regarding problems with a direct connection with this paper, it could be the first subject to extend the analysis to a model which includes congestion effects caused by overseas traffic originating on foreign networks with relation to a settlement problem of BPs with their overseas partners. In so doing, it will be necessary to consider a principle of "receiver-pays" instead of, or along with, a principle of "sender-pays" which this paper implicitly assumed, when we see today's proliferation of the use of WWW in the public. The problem of covering the costs of transmission infrastructure remains an important problem though this paper dealt with the marginal costs related to the conditions of capacity expansion.

Concerning Internet pricing, it can be said at least theoretically that "the problem of the commons" of network congestion due to "over-grazing" by users is generally solved in compliance with economic choice of entities concerned by introducing a scheme to induce users to internalize the effect of their use of network, i.e., externality of congestion, even in the decentralized and layered structure of today's real networks. This was the main result of this paper. The pricing scheme of communication considered in this paper belongs to a sort of "usage-sensitive pricing" in terms of McKnight and Baily(11), and we have focused our attention on the part of their definition of the pricing scheme, the "non-zero marginal monetary cost of sending another bit" in relation to efficient use of network resources. For the pricing scheme of the Internet, "flat-rate" has been widely used in connection with the special process of development in the U.S., and these years, offers of that scheme by providers are more frequently seen in Japan as the market becomes more competitive among them. It is claimed that the flat-rate pricing has an advantages over

other schemes in that users have no uncertainty in their communication budgets and that it enables providers to reduce costs of collecting fees on account of its simple scheme. Some people concerned with technological aspects of the Internet are said to maintain that it should be most suitable scheme for the development of the so-called "Internet culture" in view of the growing process and the technological features of the Internet. Costs of fee collection, in particular, becomes one of the important issues with which we should deal in the theoretical analysis of communication pricing as well as collecting methods in connection with the future technological developments such as improvement in protocols and inventions of new applications. The development of affairs may affect the theoretical advantage of congestion pricing. Furthermore, accompanied with a growing variety of communication contents and methods in the future, "transaction-based pricing" instead of usage-based pricing will be introduced in the market, and it gives us a problem of comparative analysis among these schemes for a subject of future studies.

In the comparative studies among pricing schemes, the results will depend on what sort of new technologies are going to be developed in the actual use of the Internet, and hence new facets and lines of analysis will be required for the economic studies on pricing and use of network resources as such new applications are actually introduced in the real world. Almost all the economic studies on information networks made so far including the present paper are mainly concerned with the analysis of "point-to-point" communication. From now on, larger efforts must be addressed to the problem of multicast flow of information; a data originates on a network domain and is replicated every time when it travels through other network nodes, forming a tree-and-branch structure of information in layered networks. Reflecting a natural line of development in Internet technology in the near future, the problem of efficient use of resources with regard to ATM has already been placed on the research agenda(12).

REFERENCES


