경영정보학연구
제7권 3호
1997년 12월

An Optimal Incentive-Compatible Pricing for Congestible Networks

김 용 재*

혼잡이 있는 네트워크를 위한 동기 유발 가격

Pricing information services, where congestion can threaten the efficient operation of information systems, has been studied in economics and information systems literature. Recent explosion of the Internet and proliferation of multimedia content over the Internet have rekindled the research interest in designing pricing schedules for differentiated information services. In order for the information system to effectively serve users having heterogeneous needs, pricing rules for discriminated services should be considered. At the same time, when individual users' interest does not align with that of the organization that individual users belong to, organization-wide pricing policy should be devised to improve the value of the services rendered by the system. This paper, using a priority queuing model, addresses the need for such an incentive-compatible pricing for different information services.

* 건국대학교 경영정보학과 조교수

이 논문은 1997년 10월 22일 접수하여 1차 수정을 거쳐 1998년 3월 3일 게재확정되었습니다.

Ⅰ. Introduction

There are few technological innovation and success stories as dramatic as that of Internet. The numbers subscribing Internet services in the US alone double almost every half a year and the speed of the Internet backbone links have increased from 56kbps to 45Mbps in just a decade. While it was initially built to link research institutions, the Internet has grown to be a social phenomenon much to the surprise of founding fathers of the network. Internet telephony, Internet-based distance learning, and Web-TV broadcasting, to name a few, are only to illustrate the tremendous growing potential of the Internet. Because the Internet, since its inception, has been functioning as an affordable yet successful testbed for new technological innovations, the range of applications that can fit into the current Internet paradigm seems to expand almost unbounded.

The Internet offers a single class of "best-effort" service; that is, there is no admission control and the network offers no assurance as to how soon and safely packets will be delivered through the network. Initially, Internet services included electronic mail (E-mail), file transfer (FTP), and name service (DNS) and yet they can tolerate certain level of delays and losses. In the presence of congestion, they can also control the transmission rate to secure reliable transmission.[1]

However, with the advent of World Wide Web (WWW) browsers, which contributed to the extreme popularity of and swift commercialization of the Internet, multimedia has been taking more of presence in the Internet and corporate Intranets. Together with commercial desires to be Web-present, graphics-intensive Web traffic becomes the largest form of Internet activity in terms of bits transferred [Varian, 1996].

Such traffic is often bursty or jittery and requires more stringent requirements in terms of latency and transmission loss. Thus [Shenker, 1995; Shenker et al., 1996; Clark, 1995] have argued that the best-effort service is too obsolete to handle the traffic having heterogeneous characteristics and it should be replaced by priority-based services in order to meet the heterogeneous demands, in particular, to suit the delay-sensitive traffic. To do that, Internet Engineering Task Force (IETF) and IEEE are working on QoS standards for the better management of the future Internet traffic and they are based on the basis of users' heterogeneous demands that they placed on the network[Shenker, 1997]. Weighted fair queuing (WFQ), dynamic bidding for access, guaranteed minimum capacity service like reservation setup protocol (RSVP) represent the concentric research efforts for preferential treatment of versatile network traffic [e.g., Schwantag, 1997; Shenker, 1996 ; Zhang, 1997].

This paper addresses the issue of allocating the bandwidth in a mission-critical corporate

1) TCP/IP protocol sends a few packets to decide the maximum transmission rate before full transmission starts. If a network becomes congested, TCP/IP recognizes the problem by receiving dropped packets before it slows down the transmission. In other words, TCP/IP tries to utilize the full capacity of the network whenever possible.

Intranet or the Internet, when there are multiple classes of users with heterogeneous demands. Following [Clark, 1995], the central hypothesis of this paper is that the Internet or Intranet services are most valued by the overall throughput during the transfer of a data object of some size. One salient aspect of this approach is that a network manager can provide differentiated services for different traffic characteristics without having the full information on individual data objects and without performing a significant overhaul of underlying network architecture. As the second-degree discrimination of economics suggests, users are offered a menu of QoS in terms of various service time distributions based on their personal preferences [Tirole, 1988; Kim, 1996; Wilson, 1989].[2)]

The plan of this paper is as follows. First, a review on the past research will show in what context this paper is positioned before we introduce a non-priority *M/G/1* system and offer an optimal and incentive-compatible pricing scheme. Second, we propose an optimal and incentive-compatible pricing scheme for nonpreemptive *M/G/1*. Third, we discuss the implications and future extension of this paper in terms of better network management.

Ⅱ. Past Research

Measuring performance of a computer system using queuing models has a long tradition. The operating characteristics of such models include mean waiting time, mean sojourn time, and the number of users served within a given time frame. From the managerial point of view, such measures, although observable to the network system manager, do not guarantee efficient control of system resources because the system value is garnered from individual users whose decentralized decision is affected by the network system manager's policy. Therefore, without explaining individual users' decision making process with microecono- mic foundations, any managerial decision cannot be comfortably defended.

Although [Pigou, 1920] was the first to point out the congestion externality of a service facility and proposed Pigouvian tax to solve the externality problem, it is [Naor, 1967] who presented an *M/M/1* system where each homogeneous user makes a decision as to joining the system or not: the aggregate effect of individual user's decentralized decisions is an equilibrium which is more congested than optimal. Thus, a fixed charge should be imposed to induce the optimal arrival rate. After Noar's seminal work, there have been numerous extensions of the model into *M/M/s* , *G/M/s*, *M/G/s*, and *M/G/1*. [Yechiali, 1971,1972; Knudsen, 1972; Mendelson, 1985] are such works.

Concurrent with the development of socially optimal solutions for various queuing paradigms, the study of economics of information raised incentive-compatibility problems. In the context of efficient management of an information system, the issue boils down to whether the socially optimal arrival rate is incentive-compatible so that decentralized decision of individual users

2) Because there is no analytical solution to general queuing networks, the solution given here can be used as an approximate solution as well as a baseline benchmark.

corresponds with the manager's socially optimal guideline.

The primary contribution of [Mendelson and Whang, 1990] is to devise an optimal incentive-compatible *Priority-and Time-Dependent* (PTD) pricing scheme for a nonpreemptive *M/M/1* system offering N priority classes for *N* user classes distinguished by *N* heterogeneous service time distributions.3) Departing from the previous queuing models hiring full information assumption on the parameter values such as the first and the second moments of service time distributions and delay costs per unit time, their work showed that the PTD pricing for nonpreemptive *M/M/1* is optimal and incentive-compatible in the sense that the arrival rates and decentralized priority selection of individual users jointly maximize the expected net value of the system.

This paper extends the idea of optimal and incentive-compatible pricing schemes of [Mendelson and Whang, 1990] to non-priority *M/G/1* and nonpreemptive priority *M/G/1* respectively in the realm of network management.4)

We state a set of assumptions similar to those in [Mendelson and Whang, 1990].

(A1) There are *N* user classes. Arrival of job s5) to the system is governed by *N* independent Poisson processes, where class-i jobs arrive at rate λ_i. We denote a priority queue serving N user classes as *M/G/1/P*=R where *M, G,* and *R* represent Markovian arrival process, general service time distributions, and the number of priority classes ($R=1,\ldots,N$) in the system, respectively.6) The class with smaller index has higher priority. We denote the non-priority system by *M/G/1/P*=1.

(A2) $V_i(\lambda_i)$ denotes the contribution of class-i jobs to system value when the class's arrival rate to the system is λ_i. $V_i(\lambda_i)$ is monotone increasing, continuously differen- tiable and strictly concave. Let $\underset{\sim}{\lambda}=(\lambda_1,\lambda_2,\ldots,\lambda_N)$ denote the arrival rate vector. The social value function is the sum of $V_i(\lambda_i)$ over all classes; i.e., $V(\underset{\sim}{\lambda})=\sum_{i=1}^{N} V_i(\lambda_i)$.7)

(A3) The service facility uses the head-of-the-line service discipline. Service time distributions are heterogeneous, and the service time required by a class-i user is generally distributed with mean c_i and second moment $c_i(2)$,

3) We distinguish the notion of user classes from that of priority classes. For example, two user classes can be served as one priority class. If a system with *N* user classes is not prioritized, it is a non-priority system.

4) In the epilogue, [Mendelson and Whang, 1990] stated that "Clearly, the next step that is called for by our results is to study the issues in a more *general* framework. ... We hope that a similar analysis can be extended to other models of interest, and particularly, to systems with other queue disciplines or a more general queuing structure." This paper attempts to do just that.

5) We use the term *user* and *job* interchangeably.

6) Therefore the number of user classes, *N*, is equal to or smaller than the number of priority classes, *R*, in the system.

7) Please see [Mendelson and Whang, 1990] for further justification of this kind of value functions.

where $c_i \neq c_j$ and $c_i^{(2)} \neq c_j^{(2)}$ if $i \neq j$. The delay cost, or sojourn time cost, per unit time for a class-i user is v_i.

(A4) The system is unsaturated, i.e., $\sum_{i=1}^{N} \lambda_i c_i < 1$, and all user classes are served in the steady state.

(A5) The network system manager is aware of the aggregate usage and cost structures for each class, while specific job characteristics are known only to the user.

(A6) We assume an individual or atomistic decision structure; i.e., users do not collude, and each will select the priority which minimizes the sum of expected access charge and sojourn time cost. We denote the sum of two costs as *total private cost* (TPC).
(See Appendix 1 for a summary of notations and definitions used in this paper.)

3. Optimal Incentive- Compatible Pricing of an *M/G/1/P*=1 System

In this section, we assume that the service facility is an $M/G/1/P$=1 system with N classes.[8] Then the mean waiting time, $W_q(\underline{\lambda})$, of a class-i job satisfies the equation

$$W_q(\underline{\lambda}) = 1/2 \sum_{k=1}^{N} \lambda_k c_k^{(2)} + \sum_{k=1}^{N} W_q(\underline{\lambda}) \lambda_k c_k = \wedge_N + W_q(\underline{\lambda}) S_N$$

where $\wedge_N = 1/2 \sum_{k=1}^{N} \lambda_k c_k^{(2)}$ and $\overline{S}_m = 1 - \sum_{k=1}^{m} \lambda_k c_k$. (1)

8) In other words, there are N different classes in the system, but none of them are prioritized. As a result, there is only one class in the system.

Solving (1) for $W_q(\underline{\lambda})$, we obtain

$$W_q(\underline{\lambda}) = \wedge_N / \overline{S}_N \ .^{9)}$$

The expected sojourn time of class-i, $ST_i^1(\underline{\lambda})$, and its derivative with respect to λ_i are[10]

$$ST_i^1(\underline{\lambda}) = W_q(\underline{\lambda}) + c_i = \wedge_N / \overline{S}_N + c_i \quad \text{and} \qquad (2)$$

$$\frac{\partial ST_i^1(\underline{\lambda})}{\partial \lambda_j} = \frac{c_j^{(2)}}{2\overline{S}_N} + \frac{\wedge_N c_j}{\overline{S}_N^2}$$

respectively. The total delay cost, $TC^1(\overline{\lambda})$, of $M/G/1/P$=1 is

$$TC^1(\underline{\lambda}) = \sum_{j=1}^{N} v_j \lambda_j ST_j^1(\underline{\lambda}). \qquad (3)$$

The net-system-value-maximizing problem is stated as

$$\max_{\overline{\lambda}} V(\underline{\lambda}) - TC^1(\underline{\lambda}) = \max_{\overline{\lambda}} \sum_{j=1}^{N} (V_j(\lambda_j) - v_j \lambda_j (\wedge_N / \overline{S}_N + c_j)) \qquad (4)$$

and the first-order conditions for optimality are, for $i = 1, \ldots, N$,

$$V_i(\lambda_i) = v_i ST_i^1(\underline{\lambda}) + \sum_{j=1}^{N} v_j \lambda_j \frac{\partial TC^1(\underline{\lambda})}{\partial \lambda_i}$$

$$= v_i ST_i^1(\underline{\lambda}) + \left(\frac{c_i^{(2)}}{2\overline{S}_N} + \frac{\wedge_N c_i}{\overline{S}_N^2} \right) \sum_{j=1}^{N} v_j \lambda_j \qquad (5)$$

which simply states that the social marginal value should be equal to the social marginal delay cost at an optimal arrival vector $\underline{\lambda}^*$. Suppose that there exists such an optimal

9) $\wedge_N$ represents mean delay time caused by the job being served when the tagged job arrives at the system, and $W_q(\underline{\lambda}) S_N$ is the additional mean delay time caused by the jobs ahead of the tagged job.

10) The superscript "1" denotes that there is one class in the system.

solution[11], $\underline{\lambda}^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_N^*)$, for (5) and that the system manager announces the optimal access charge for class-i users, $P_i^1(\underline{\lambda}^*)$, which is equal to the externality cost of (5) at $\underline{\lambda}^*$.[12]
In other words,

$$P_i^1(\underline{\lambda}^*) = \left(\frac{c_i^{(2)}}{2\bar{S}_N^*} + \frac{\Lambda_N c_i}{\bar{S}_N^{*2}}\right)\sum_{j=1}^{N} v_j\lambda_j^* \quad (i=1,2,\ldots,N) \quad (6)$$

where $\bar{S}_m^* = 1 - \sum_{k=1}^{m} \lambda_k^* c_k$.

The term in parentheses is dependent on c_i and $c_i^{(2)}$, and is the source of the incentive-compatibility problem because the manager cannot tell which class a given arriving job belongs to. That is, a class-i user who is supposed to pay $P_i^1(\underline{\lambda}^*)$ will not pay $P_i^1(\underline{\lambda}^*)$. Rather, in order to save the access charge cost, he will select $P_m^1(\underline{\lambda}^*)$, the minimum of all $P_j^1(\underline{\lambda}^*)$'s $(j=1,\ldots,N)$, contrary to the manager's intention.

Example 3.1.

Suppose that there are two user classes in an M/M/1 system where, $V_1(\lambda_1) = 9\lambda_1 - 20\lambda_1^2$, $V_2(\lambda_2) = 12\lambda_2 - 30\lambda_2{}^2$, $v_1 = 2$, $v_1 = 1$, $c_1 = 0.1$ and $c_2 = 2$. Solving the first-order conditions (5) gives the optimal traffic vector $(\lambda_1^*, \lambda_2^*) = (0.375, 0.111)$ and the optimal price vector $\underline{P}^* = (P_1^*, P_2^*) = (0.082, 6.064)$. With $(\lambda_1^*, \lambda_2^*)$, the net system value is 2.298. However, if $\underline{P}^*$ is announced, class-2 users will take price P_1^* over P_2^* because it is advantageous to do so. If all the class-2 users select P_1^*, the system traffic will be $(\lambda_1^+, \lambda_2^+) = (0.215, 0.257)$ and the net system value 1.467, which is a 35% reduction from the optimal state but almost no improvement over the non-intervention case. Lack of intervention will make the net system value equal to 1.467 and the arrival rates will be $(\lambda_1, \lambda_2) = (0.215, 0.258)$. ||

The above example shows that optimality and incentive-compatibility issues should be dealt separately: if the network manager wants to maintain an optimal system traffic level, he also should make the pricing incentive- compatible. Otherwise, his goal of maximizing the net system value achieved from the network services can be jeopardized as the previous example illustrates.

The incentive problem stems from the heterogeneity of service time distributions and yet identical expected waiting times time spent before service begins across all classes: a class-i user will choose the lowest access charge, $P_m^1(\underline{\lambda}^*)$, while enjoying the same waiting time. In other words, as pointed out by [Mendelson and Whang, 1990], $P_i^1(\underline{\lambda}^*)$ is not incentive-compatible. It is therefore necessary for the network system manager to devise a *time-dependent* (TD) pricing scheme, which is incentive-compatible and optimal.[13] TD pricing was first conceived by [Mendelson

11) Because our analysis is based on *Nash*-equilibrium, multiple *Nash*-equilibria are likely. Following [Mendelson and Whang, 1990], we assume that if there is more than one solution, only the solution of the smallest magnitude is selected.
12) See [Mendelson and Whang, 1990] for the detailed derivation of the terms used in this paper.
13) [Shenker, 1996; Clark, 1995] argued that usage-based pricing should be a rule in the future Internet pricing.

and Whang, 1990], but we underscore the fact that the incentive-compatible pricing should be employed not only for priority queues but also for a non-priority queue if service time distributions are heterogen- eous.

THEOREM 3.1.

Suppose the manager announces a single TD price for all jobs in the non-priority M/G/1/P=1 system such that

$$p^1(t)=\left(\frac{t^2}{2\overline{S}_N^*}+\frac{\wedge_N t}{\overline{S}_N^{*2}}\right)\sum_{j=1}^{N} v_j\lambda_j^*. \quad (7)$$

Then $P^1(t)$ *is optimal and incentive-compatible.*[14]

PROOF:

Let E_i be an expectation operator on class-i service time t. I.e., $E_i[t^m]=c_i^{(m)}$.

Then $E_i[p^1(t)]=(c_i^{(2)}/2\overline{S}_N^*+\wedge_N c_i/\overline{S}_n^{*2})\sum_{k=1}^{N} v_k\lambda_k$ $=p_i^1(\lambda^*)$ for $i=1,\ldots,N$. Thus $P^1(t)$ is optimal. It is also incentive-compatible because the usage-based price does not provide any room for personal arbitrage, i.e., switching classes. ||

Example 3.2.

We examine the incentive problem as in Example 3.1 but with the TD pricing. From (7), we obtain $p^1(t)=1.164t^2+0.704t$ which will force the class-2 users to reveal his true usage: class-2 users will pay 6.064 as is required by the optimal condition in (6) while class-1 users will pay 0.082. Consequently, the system will reach the optimal state with the net system value 2.298. ||

Before we move to the next section, we state the well known optimal priority assignment rule which minimizes the total delay costs of M/G/1/P=N and M/M/1/P=N.

THEOREM 3.2.

For nonpreemptive M/G/1/P=N and preemptive-resume M/M/1/P=N, the v_i/c_i *rule* $(v_1/c_1\geq v_2/c_2\geq\ldots\geq v_N/c_N)$ *is the optimal priority assignment for feasible* $(\lambda_1,\lambda_2,\ldots,\lambda_N)$.

Proof:

See [Jaiswal, 1968] and [Mendelson and Whang, 1990]. ||

In other words, the v_i/c_i rule is the optimal priority assignment policy for nonpreemptive M/G/1/P=N and preemptive M/M/1/P=N, regardless of changes in the system traffic.[15]

4. Optimal Incentive- Compatible Pricing For Nonpreemptive Priority M/G/1/P=N

This section deals with a usage-based pricing in nonpreemptive M/G/1/P=N queue.

---

14) We assume that users' preference is risk-neutral in this paper.

15) Although the rule can be applied to a specific case of preemptive M/G/1/P=N, proving the optimality of v_i/c_i rule for general M/G/1/P=N is not possible. Also note that the "preemptive" assumption may not make sense for the Internet or Intranets where the networked computing systems are often too loosely connected to preempt lower-class jobs in the middle of transmission.

Suppose that the service facility is run as nonpreemptive $M/G/1/P{=}N$ and let $ST_i(\underline{\lambda})$ and $TC(\underline{\lambda})$ denote the expected sojourn time of class-i and the total delay cost of nonpreemptive $M/G/1/P{=}N$ respectively. Then $ST_i(\underline{\lambda}) = \wedge_N / \overline{S}_i \overline{S}_{i-1} + c_i$ and $TC(\underline{\lambda}) = \sum_{j=1}^{N} v_j \lambda_j ST_j(\underline{\lambda})$ (See[Kleinrock, 1976]).

The net system value is given by the expression

$$\sum_{i=1}^{N} V_i(\lambda_i) - \sum_{i=1}^{N} v_i \lambda_i ST_i(\underline{\lambda})$$
$$= \sum_{i=1}^{N} V_i(\lambda_i) - \sum_{i=1}^{N} v_i \lambda_i (\wedge_N / \overline{S}_{i-1} \overline{S}_i + c_i), \quad (8)$$

and the first-order conditions are, for $i = 1, 2, \ldots, N$

$$V_i'(\lambda_i) = v_i ST_i(\underline{\lambda}) + \sum_{j=1}^{N} v_j \lambda_j \left(\frac{c_i^{(2)}}{2\overline{S}_{j-1}\overline{S}_j} + \frac{1_{\{j-1 \geq i\}} c_i \wedge_N}{\overline{S}_{j-1}^2 \overline{S}_j} + \frac{1_{\{j \geq i\}} c_i \wedge_N}{\overline{S}_{j-1} \overline{S}_j^2} \right) \quad (9)$$

where $1_{\{Cond\}}$ represents the indicator function. That is, $1_{\{Cond\}} = 1$ if *Cond* is true and 0 otherwise.

Suppose that (9) is solved and its optimal arrival rate vector is $\underline{\lambda}^*$. Then the optimal access charge for class i, $P_i(\underline{\lambda}^*)$, is

$$P_i(\underline{\lambda}^*) = \sum_{j=1}^{N} v_j \lambda_j \left(\frac{c_j^{(2)}}{2\overline{S}_{j-1}^* \overline{S}_j^*} + \frac{1_{\{j-1 \geq i\}} c_i \wedge_N^*}{\overline{S}_{j-1}^{*2} \overline{S}_j^*} + \frac{1_{\{j \geq i\}} c_i \wedge_N^*}{\overline{S}_{j-1}^* \overline{S}_j^{*2}} \right) \quad (i = 1, 2, \ldots, N) \quad (10)$$

where

$$\overline{S}_j^* = 1 - \sum_{k=1}^{j} \lambda_k^* c_k \text{ and } \wedge_j^* = 1/2 \left(1 - \sum_{k=1}^{j} \lambda_k^* c_k^{(2)}\right).$$

Again $P_i(\underline{\lambda}^*)$ is not incentive-compatible: a class-i user will select class-j priority if

$$p_i(\underline{\lambda}^*) + v_i ST_i(\underline{\lambda}^*) >$$
$$p_j(\underline{\lambda}^*) + v_i (ST_j(\underline{\lambda}^*) - c_j + c_i) \quad (11)$$

The inequality (11) indicates that a class-i user will renege class-i priority queue if doing so can make him better off. Unless the left-hand-side of (11), the TPC from selecting the class-i priority, is smaller than that from opting class-j priority (the right- hand-side of the inequality), class-i users will select the class-j priority. The next example shows that lack of incentive-compatibility conditions can lead the system into a sub-optimal state due to class-i users' action against the manager's intention.

Example 4.1.

Suppose that there are two user classes in an $M/M/1$ system where $V_1(\lambda_1) = 9\lambda_1 - 20\lambda_1^2$, $V_2(\lambda_2) = 12\lambda_2 - 30\lambda_2^2$, $v_1 = 2$, $v_1 = 1$, $c_1 = 0.1$ and $c_2 = 2$ as in Example 3.1. Solving the first-order conditions (9) give the optimal arrival rate vector $(\lambda_1^*, \lambda_2^*) = (0.372, 0.153)$ and the optimal pricing vector of (10) is $(P_1^*, P_2^*) = (0.081, 4.462)$. For $(\lambda_1^*, \lambda_2^*)$, the net system value of (8) is 2.449. Table 1 summarizes the total private costs when class-i users selects class-j priority at $(\lambda_1^*, \lambda_2^*)$. Suppose that the manager announces $(p_1^*, p_2^*) = (0.081, 4.462)$ for class-1 and class-2 users respectively. Given the priority and access price choice, class-2 users will select class-1 priority access charge 0.081 instead of paying 4.462 because his TPC in (11) will be reduced from 7.402 to 2.722 as is shown in

Table 1. If all class-2 users declare to be class-1, the system is run as a non-priority system and all users are paying the access charge 0.081.

<Table 1> The total private cost when

$(\lambda_1^*, \lambda_2^*) = (0.372, 0.153)$

Class \ Priority	1	2
1	1.563	6.542
2	2.722	7.402

With all class-2 users selecting class-1 priority, the optimal arrival rate $(\lambda_1^*, \lambda_2^*)$ cannot be sustained: the priority designated for class-2 users becomes extinct and the problem boils down to solving the non-priority *M/M/1*. The new equilibrium arrival rate vector is $(\lambda_1^+, \lambda_2^+) = (0.215, 0.257)$, which shows that the arrival rate of class-2 users increases at the expense of class-1 users. Surprisingly, the net system value at $(\lambda_1^+, \lambda_2^+)$ is only 1.488, a 39% reduction from 2.449. ‖

The aforementioned example clearly illustrates the danger of implementing a solution fixing congestion externality. Not only the outcome falls off the optimal state, but also the net system value declines substantially. Therefore, in order to prohibit the personal arbitrage, the network manager will provide a *Priority-and Time-dependent* (PTD) price, $p_i(t)$, which induces a class-*i* user to reveal his true service requirement. The following theorem augments [Mendelson and Whang, 1990] by employing general service time distributions.

Theorem 4.3.

The TDP, given below as $p_i^*(t)$, *of a nonpreemptive M/G/1/P=N system is optimal and incentive-compatible where*

$$p_i^*(t) = A_i t + 1/2Bt^2 \tag{12}$$

where

$$B = \sum_{k=1}^{N} \frac{v_k \lambda_k^*}{\overline{S}_{k-1}^* \overline{S}_k^*} \quad \text{and}$$

$$A_i = \frac{v_i \lambda_i^* \wedge_N}{\overline{S}_{i-1}^* \overline{S}_i^{*2}} + \sum_{k=i+1}^{N} \left(\frac{\wedge_N v_k \lambda_k^*}{\overline{S}_k^* \overline{S}_{k-1}^{*\,2}} + \frac{\wedge_N v_k \lambda_k^*}{\overline{S}_{k-1}^* \overline{S}_k^{*2}} \right)$$

Proof: 16)

$p_i(t)$ is optimal because

$$Ei[p_i(t)] = \frac{v_i \lambda_i^* \wedge_N c_i}{\overline{S}_{i-1}^* \overline{S}_i^{*2}} + \sum_{k=i+1}^{N} V_k \lambda_k^* c_i \wedge_N \left(\frac{1}{\overline{S}_k^* \overline{S}_{k-1}^{*\,2}} + \frac{1}{\overline{S}_{k-1}^* \overline{S}_k^{*2}} \right)$$

$+1/2 \sum_{k=1}^{N} \frac{v_k \lambda_k^*}{\overline{S}_{k-1}^* \overline{S}_k^*} c_i^{(2)} = p^i(\underline{\lambda}^*)$. Our remaining job is to show that $p_i(t)$ is incentive- compatible. In order to show this, we first define total private cost (*TPC*) as the actual cost perceived by individual users. Using *TPCs*, we introduce a cheating penalty function $\Pi^i(j)$, which represents the differential of total private cost (*TPC*) when a class-*i* user selects class-*j* priority rather than class-*i* priority $(i \neq j)$.

The cheating penalty function, which represents the penalty that a class-*i* user should pay if he selects class *j*, will be

$$\Pi^i(j) = E_i[p_j(t)] + v_i(ST_j(\underline{\lambda}^*) - c_j + c_i) - E_i[p_i(t)] - v_i ST_i(\underline{\lambda}^*)$$

$$= A_j c_i - A_i c_i + v_i \wedge_N (1/\overline{S}_{j-1}^* \overline{S}_i^* - 1/\overline{S}_{i-1}^* \overline{S}_i^*)$$

(See [Mendelson and Whang, 1990]).

Obviously $\Pi^i(i) = 0$. The PTD pricing

16) For an intuitive explanation of why PTD pricing is quadratic in *t*, refer to [Mendelson and Whang, 1990].

vector $(p_1(t), p_2(t), \ldots, p_N(t))$ is incentive-compatible if and only if $(\Pi^i(j) > 0 \ (i \neq j))$.

We prove the incentive-compatibility by proving a stronger claim such that $\Pi^i(k)$ is a unimodal function of k and the minimum occurs when $k = i$.

First, for $i < j$,

$$\Pi^i(j) - \Pi^i(j+1)$$

$$= \wedge_N\left(\frac{v_j\lambda_j^* c_i}{\bar S_{j-1}^* \bar S_j^{*2}} - \frac{v_{j+1}\lambda_{j+1}^* c_i}{\bar S_j^* \bar S_{j+1}^{*\,2}} + \left(\frac{v_{j+1}\lambda_{j+1}^* c_i}{\bar S_{j+1}^* \bar S_j^*} + \frac{v_{j+1}\lambda_{j+1}^* c_i}{\bar S_j^* \bar S_{j+1}^{*\,2}}\right)\right.$$
$$\left. + \left(\frac{v_i}{\bar S_{j-1}^* \bar S_j^*} - \frac{v_i}{\bar S_j^* \bar S_{j+1}^*}\right)\right)$$

$$= \wedge_N\left(\frac{v_j\lambda_j^* c_i}{\bar S_{j-1}^* \bar S_j^{*2}} + \frac{v_{j+1}\lambda_{j+1}^* c_i}{\bar S_{j+1}^* \bar S_j^{*2}} - \frac{v_i(\lambda_{j+1}^* c_{j+1} + \lambda_j^* c_j)}{\bar S_{j-1}^* \bar S_j^* \bar S_{j+1}^*}\right)$$

$$< \wedge_N\left(\frac{v_j\lambda_j^* c_i}{\bar S_{j-1}^* \bar S_j^{*2}} + \frac{v_{j+1}\lambda_{j+1}^* c_i}{\bar S_{j+1}^* \bar S_j^{*2}} - \frac{v_{j+1}\lambda_{j+1}^* c_i + v_j\lambda_j^* c_i}{\bar S_{j-1}^* \bar S_j^* \bar S_{j+1}^*}\right)$$

(because $v_i c_{j+1} > v_{j+1} c_i$ and $v_i c_j > v_j c_i$)

$$= \wedge_N\left(\frac{v_j\lambda_j^* c_i(\bar S_{j+1}^* - \bar S_j^*)}{\bar S_{j-1}^* \bar S_j^{*2} \bar S_{j+1}^*} + \frac{v_{j+1}\lambda_{j+1}^* c_i}{\bar S_{j+1}^* \bar S_j^{*2}} - \frac{v_{j+1}\lambda_{j+1}^* c_i}{\bar S_{j-1}^* \bar S_j^* \bar S_{j+1}^*}\right)$$

$$= \wedge_N\left(-\frac{v_j\lambda_j^* c_i\lambda_{j+1}^* c_{j+1}}{\bar S_{j-1}^* \bar S_j^{*2} \bar S_{j+1}^*} + \frac{v_{j+1}\lambda_{j+1}^* c_i}{\bar S_{j+1}^* \bar S_j^{*2}} - \frac{v_{j+1}\lambda_{j+1}^* c_i}{\bar S_{j-1}^* \bar S_j^* \bar S_{j+1}^*}\right)$$

$$< \wedge_N\left(-\frac{v_{j+1}\lambda_j^* c_i\lambda_{j+1}^* c_j}{\bar S_{j-1}^* \bar S_j^{*2} \bar S_{j+1}^*} + \frac{v_{j+1}\lambda_{j+1}^* c_i}{\bar S_{j+1}^* \bar S_j^{*2}} - \frac{v_{j+1}\lambda_{j+1}^* c_i}{\bar S_{j-1}^* \bar S_j^* \bar S_{j+1}^*}\right)$$

(because $v_j c_{j+1} > v_{j+1} c_j$)

$$= \wedge_N V_{j+1}\lambda_{j+1}^* C_i\left(\frac{1}{\bar S_{j-1}^* \bar S_j^* \bar S_{j-1}^*} - \frac{1}{\bar S_{j-1}^* \bar S_j^* \bar S_{j+1}^*}\right) = 0$$

Therefore $\Pi^i(i) - \Pi^i(i+1) < \Pi^i(i+2) < \ldots < \Pi^i(N)$.

The other half of the proof is for $i > j$.

$$\Pi^i(j) - \Pi^i(j-1)$$

$$= \wedge_N\left(\frac{v_j\lambda_j^* c_i}{\bar S_{j-1}^* \bar S_j^{*2}} - \frac{v_{j-1}\lambda_{j-1}^* c_i}{\bar S_{j-2}^* \bar S_{j-1}^{*\,2}} - \left(\frac{v_j\lambda_j^* c_i}{\bar S_j^* \bar S_{j-1}^{*\,2}} + \frac{v_j\lambda_j^* c_i}{\bar S_{j-1}^* \bar S_j^{*2}}\right)\right.$$
$$\left. + \left(\frac{v_i}{\bar S_{j-1}^* \bar S_j^*} - \frac{v_i}{\bar S_{j-1}^* \bar S_{j-2}^*}\right)\right)$$

$$= -\frac{v_{j-1}\lambda_{j-1}^* c_i \wedge_N}{\bar S_{j-2}^* \bar S_{j-1}^{*\,2}} - \frac{v_j\lambda_j^* c_i \wedge_N}{\bar S_{j-1}^{*\,2} \bar S_j^*} + \frac{v_i \wedge_N}{\bar S_{j-1}^* \bar S_j^*} - \frac{v_i \wedge_N}{\bar S_{j-2}^* \bar S_{j-1}^*}$$

$$< -\frac{v_i\lambda_{j-1}^* c_{j-1} \wedge_N}{\bar S_{j-2}^* \bar S_{j-1}^{*\,2}} - \frac{v_i\lambda_j^* c_j \wedge_N}{\bar S_{j-1}^{*\,2} \bar S_j^*} + \frac{v_i \wedge_N}{\bar S_{j-1}^* \bar S_j^*} - \frac{v_i \wedge_N}{\bar S_{j-2}^* \bar S_{j-1}^*}$$

(because $v_i c_{j-1} < v_{j-1} c_i$ and $v_i c_j < v_j c_i$)

$$= v_i \wedge_N\left(-\frac{\lambda_{j-1}^* c_{j-1}\bar S_j^* + \lambda_j^* C_j \bar S_{j-2}^*}{\bar S_{j-2}^* \bar S_{j-1}^{*\,2} \bar S_j^*} + \frac{1}{\bar S_{j-1}^* \bar S_j^*} - \frac{1}{\bar S_{j-2}^* \bar S_{j-1}^*}\right)$$

$$= v_j \wedge_N\left(-\frac{\lambda_{j-1}^* c_{j-1}(\bar S_{j-1}^* - \lambda_j^* c_j) + \lambda_j^* c_j(\bar S_{j-1}^* + \lambda_{j-1}^* c_{j-1})}{\bar S_{j-2}^* \bar S_{j-1}^{*\,2} \bar S_j^*}\right.$$
$$\left. + \frac{1}{\bar S_{j-1}^* \bar S_j^*} - \frac{1}{\bar S_{j-2}^* \bar S_{j-1}^*}\right) = 0$$

Thus $\Pi^i(1) > \Pi^i(2) > \ldots > \Pi^i(j) > \Pi^i(j-1) > \ldots > \Pi^i(i) = 0$.
Because $\Pi^i(k)$ is a unimodal function and its minimum 0 occurs at $k = i$, there is no incentive for a class-i user to select class-j priority over class-i priority. ||

Example 4.2.

Suppose that there are two user classes in an M/M/1 priority system where $V_1(\lambda_1) = 9\lambda_1 - 20\lambda_1^{\,2}$, $V_2(\lambda_2) = 12\lambda_2 - 30\lambda_2^{\,2}$, $v_1 = 2$, $v_1 = 1$, $c_1 = 0.1$ and $c_2 = 2$ as in Example 3.1. Solution for the first-order condition (9) is $(\lambda_1^*, \lambda_2^*) = (0.372, 0.153)$and the PTD pricing of (12)is $(p_1^*(t), P_2^*(t)) = (1.006t^2 + 0.714t, 1.006t^2 + 0.219t)$

The net system value is 2.449 in this case, which is a 6% improvement over the net system value of non-priority M/M/1. ||

5. Concluding Remarks

In this paper, we discussed incentive-compatible pricing schemes a variation of Pigouvian tax. Both issues should be addressed before the system manager can reach an optimal solution for non-priority M/G/1 and nonpreemp- tive priority M/G/1 for network management. We argued that incentive- compatible pricing is necessary for optimal output even under non-priority M/G/1. We also augmented the PTD pricing scheme of [Mendelson and Whang, 1990] for nonpreemptive M/G/1. We summarize the conclusion of this paper in Table 2.

<Table 2> v_i/c_i Rule, IC Constraints, and Queuing Discipline.

Queuing Discipline	Optimality of the v_i/c_i rule	Incentive-compatibility of PTD pricing
Nonpreemptive M/G/1/P=N	Yes. Can prove analytically	Yes. Can prove analytically
Preemptive-resume M/M/1/P=N	Yes. Can prove analytically [Mendelson and Whang 90]	Yes. Can prove analytically [Mendelson and Whang 90]

There are at least a couple of problems in applying the priority-pricing scheme of this paper to congestion-prone networks. First, it can be asked which data object (job) should be given a preferential treatment over another. For example, should e-mail message be given a lower priority over video message over the Internet or corporate Intranets? Although conventional wisdom taught that video message, which is more susceptible to transmission delay, should be given a higher priority over text-based traffic, there may be very urgent mail massages to get through the network.

In such a case, the manager should provide a means to let the urgent messages grab the immediate attention of the network; that is, identifying such requests should not be based on QoS, application types, data size, data path, or an IP address, but should be based on the individual preferences and valuation of the transferred data object.

Second, although this paper assumed that the number of priority classes are given *a priori*, the network manager should ask how many different priority classes can be provided and how he should segment the whole user population into multiple classes.

Apparently, the pursuit of these questions opens a wide avenue of research because analysis of user demand patterns, the congestion costs perceived by individual users, and individual valuation of services should be preceded in order to make our model close to reality. [Kim, 1996] proposed a method partitioning user population into N classes and [Wilson, 1989] discussed the elfare implications of classification efforts.

〈REFERENCES〉

Alperstein, H., "Optimal Pricing Policy for the Service Facility Offering a Set of Priority Prices," *Management Science*, 34, May 1988, pp. 666-671.

Balachandran, K. R. and M. E. Shaefer, "Class Dominance Characteristics at a Service Facility," *Management Science*, Vol.47, March 1979, pp. 515-519.

Balachandran, K. R. and S. Radhakrishnan, "Extensions to class dominance characteristics," *Management Science*, Vol.40, No.10, October 1994, pp. 353-360.

Balachandran, K. R., "Purchasing Priorities in Queues," *Management Science*, Vol.18, No.5, January 1972, pp. 316-326.

Beja, A. and E. Sid, "Optimal Priority Assignment with Heterogeneous Waiting

Costs," *Operations Research,* Vol.23, No.1, July 1975, pp. 107-117.

Bell, C. and S. Stidham, "Individual versus Social optimization in the Allocation of Customers to Alternative Servers," *Management Science,* July 1983.

Chakravorti, B., "Optimal Flow Control of an M/M/1 Queue with a Balanced Budget," *IEEE Transactions on Automatic Control,* Vol.39, No.9, September 1994, pp. 1918-1921.

Chao, H. and R.B. Wilson, "Optimal Contract Period For Priority Service," *Operations Research,* Vol.38, No.4, 1990, pp. 598-606.

Chao, H., "Priority Service: Pricing, Investment, and Market Organization," *American Economic Review,* Vol.77, No.5, December 1987, pp. 899-916.

Clark, D. D., "Adding Service Discrimination to the Internet," *Telecommunications Policy,* Vol.20, No.3, 1996, pp. 169-181.

DeVanny, A. and G. Saving, "The Economics of Quality," *Journal of Political Economy,* 91, June 1983, pp. 979-1000.

DeVany, A. and N.G. Frey, "Stochastic Equilibrium and Capacity Utilization," *AEA Papers and Proceedings,* May 1981, pp. 53-57

DeVany, A. and T.R. Saving, "The Economics of Quality," *Journal of Political Economy,* 91, June 1983, pp. 979-1000.

DeVany, A., "Uncertainty, Waiting Time, and Capacity Utilization: A Stochastic Theory of Product Quality," *Journal of Political Economy,* Vol.84, No.3, 1976, pp. 523-541.

Dewan, S., H. Mendelson, "User Delay Costs and Internal Pricing for a Service Facility," *Management Science,* Vol.36, No.12, December 1990, pp. 1502-17.

Dolan, R., "Incentive Mechanisms for Priority Queuing Problems," *The Bell Journal of Economics* Vol.9, No.2, 1978, pp. 421-436.

Ekelund, R., "Price Discrimination and Product Differentiation in Economic Theory: An early Analysis," *Quarterly Journal of Economics,* 84, pp. 268-278.

Giridharan, P.S. and H. Mendelson, "Free-access Policy for Internal Networks," *Information Systems Research,* Vol.5, No.1, March 1994, pp. 1-21.

Gross, D. and C. Harris, *Fundamentals of Queuing Theory,* John Wiley & Sons, Inc., 1985.

Gurbaxani, V. and H. Mendelson, "An Integrative Model of Information Systems Spending Growth," *Information Systems Research,* Vol.1, No.1, March 1990, pp. 23-46.

Jaiswal, N. K., *Priority Queues,* New York, Academic Press, 1968.

Kim Y. and J. Langford, The Pareto-improving Transition to Prioritized Service for Incentive-compatible *M/G/1* queues, Department Working Paper, University of Washington, December 1995.

Kleinrock, L., "Optimum Bribing for Queue Position," *Operations Research,* Vol.15, No.2, March-April 1967, pp. 304-318.

Kleinrock, L., "Purchasing Priorities in Queues," *Management Science,* Vol.18, No.5, January 1972, pp. 319-326.

Kleinrock, L., *Queuing Systems, Vol. II: Computer Applications*, John Wiley & Sons, Inc. 1976.

Knudsen, N.C., "Individual and Social Optimization in Multi-Serve Queue with a General Cost-Benefit Structure," *Econometric,* Vol.40, No.3, 1972, pp. 515-528.

Levhari, D. and I. Luski, "Duopoly Pricing and Waiting Lines," *European Economic Review,* Vol.11, 1978, pp. 17-35.

Lippman, S.A. and Stidham, S. Jr., "Individual versus Social Optimization in Exponential Congestion Systems," *Operations Research,* Vol.25, No.2, 1977, pp. 232-247.

Lui, F.T., "An Equilibrium Queuing Model of Bribery," *Journal of Political Economy,* Vol.93, No.4, 1985, pp. 761-781.

MacKie-Mason, J.K. and H.R. Varian, Pricing the Internet, University of Michigan Working Paper, June 1993a.

MacKie-Mason, J.K. and H.R. Varian, Some Economics of the Internet, University of Michigan Working Paper, June 1993b.

Marchand, M., "Priority Pricing," *Management Science,* Vol.20, March 1974, pp. 1131-1140.

Markus, M.L., "Chargeback as an Implementation Tactic for Office Communication Systems," *Interfaces,* Vol.17, No.3, 1987, pp. 54-63.

Mendelson, H. and S. Whang, "Optimal Incentive-compatible Priority Pricing for the *M/M/1* Queue," *Operations Research,* Vol.38, No.5, Sep.-Oct. 1990, pp. 870-83.

Mendelson, H., "Economics of Scale in Computing: Grosch's Law Revisited," *Communications of ACM,* Vol.30, No.12, 1987, pp. 1066-1072.

Mendelson, H., "Pricing Computer Services: Queuing Effects," *Communications of the ACM,* Vol.28, No.3, 1985, pp. 312-321.

Miller, B.L. and A.G. Buckman, "Cost Allocation and Opportunity Costs," *Management Science,* Vol.33, No.5, 1987.

Miller, B.L., "A Queuing Reward System with Several Customer Classes," *Management Science,* Vol.16, No.3, 1969.

Murphy, L., J. Murphy and J. K. MacKie-Mason, Feedback and Efficiency in ATM Networks, Ann Arbor: University of Michigan Department of Economics, 1996.

Mussa, M. and S. Rosen, "Monopoly and Product Quality," *Journal of Economic Theory,* Vol.18, 1978, pp. 301-317.

Naor, P., "The Regulation of Queuing Size by

Levying Tolls," *Econometrica* Vol.37, No.1, January 1969, pp. 15-24.

Pigou, P., *The Economics of Welfare,* First Edition, Macmillan, London, 1920.

Shenker, S., D. D. Clark, D. L. Estrin and S. Herzog, "Pricing in Computer Networks: Reshaping the Research Agenda," *Telecommunications Policy,* Vol.20, No.3, 1996, pp. 183-201.

Shenker, S., Service Models and Pricing Policies for an Integrated Services Internet, Palo Alto: Xerox Corporation Palo Alto Research Center, 1995.

Shenker, S., Network Element Service Specification Template, Palo Alto: Xerox Corporation Palo Alto Research Center, Request for Comments:2216, Sep, 1997.

Stidham, S. Jr. and R. R. Weber, "Optimal Flow Control of an M/M/1 Queue with a Balanced Budget," *IEEE Transactions on Automatic Control,* Vol.39, No.9, September 1994, pp. 1918-1921.

Shenker, S. and J. Wroclawski, General Characterization Parameters for Integrated Service Network Elements, Palo Alto: Xerox Corporation Palo Alto Research Center, Internet-Drafts, Jul, 1997.

Stidham, S. Sr. and R. R. Weber, "A survey of Markov decision models for control of networks of queues," *Queuing Systems Theory and Applications,* Vol.13, No.3, 1993, pp. 291-314.

Stidham, S. Sr. and R. R. Weber, "Monotonics and Insensitive Optimal Policies for Control of Queues with Discounted Costs," *Operations Research,* Vol.87, No.4, 1989, pp. 611-625.

Stidham, S. and R. Weber, "A survey of Markov decision models for control of networks of queues," *Queuing Systems Theory and Applications,* Vol.13, No.3, 1993, pp. 291-314.

Stidham, S., "Pricing and Capacity Decisions for a Service Facility: Stability and Multiple Local Optima," *Management Science,* Vol.38, No.8, August 1992, pp. 1121-1139.

Takagi, H., *Queuing analysis: A foundation of performance evaluation,* New York: Elsevier, 1991.

Tirole, J., *The theory of Industrial Organization,* MIT Press, Cambridge, Mass, 1988.

Varian, H.R., Economic Incentives in Software Design, University of Michigan Working Paper, June 1993.

Varian, H.R., Economic Issues Facing the Internet, University of Michigan Working Paper, June 1996.

Whang, S., "Alternative Mechanisms of Allocating Computer Resources Under Queuing Delays," *Information Systems Research* Vol.1, No.1, March 1990, pp. 71-88.

Whang, S., Pricing Computer Systems: Incentive, Information and Queuing Effects, Ph.D. Dissertation, W.E. Simon Graduate

School of Business Administra- tion, University of Rochester, 1988.

Wilson, R. B., Economic Theories of Price Discrimination and Product Differentiation: A Survey, Technical notes, Stanford Business School, Stanford University, July 1991.

Wilson, R. B., "Efficient and Competitive Rationing," *Econometrica,* Vol.57, No.1, January 1989, pp. 1-40.

Wilson, R. B., *Non-linear pricing,* Oxford University Press, 1993.

Yechiali, U., "Customers' Optimal Joining Rules for the *GI/M/1* Queue," *Management Science,* Vol.18, July 1972, pp. 434-443.

Yechiali, U., "On Optimal Balking Rules and Toll Charges in the *GI/M/1* Queue Process," *Operations Research,* September 1971, pp. 348-370.

Zhang Z., End-to-end Support for Statistical Quality-of-Service Guarantees in Multimedia Networks, Ph.D. Dissertation, University of Minnesota, 1997.

◆ 저자소개 ◆

김 용 재 (Kim, Yong Jae)
서울대학교에서 1979년 경제학 학사, SUNY at Stony Brook에서 1983년도 경제학 석사, University of Kansas에서 1985년 전산과학 석사, University of Washington에서 1996년 경영학 박사를 취득한 후, 1996년 9월부터 건국대학교 경영정보학과에서 재직하고 있다.

Appendix 1. Summary of Notations and Definitions

N = number of classes in the system

P = number of priorities in the system

v_i = class-i delay time cost per unit time

$c_j^{(k)}$ = k-th moment of the service time distribution of a class-i job

c_i = mean of the service time distribution of a class-i job

λ_i = class-i arrival rate

$\underline{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ = system arrival rate vector

$S_i = \sum_{m=1}^{i} \lambda_m c_m$

$\overline{S}_i = 1 - S_i$

$\wedge_i = \sum_{m=1}^{i} \lambda_m c_m^{(2)}/2$

$\underline{\lambda}^+ = (\lambda_1^+, \lambda_2^+, \ldots, \lambda_N^+)$ = optimal system traffic vector of non-priority system

$\underline{\lambda}^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_N^*)$ = optimal system traffic vector of a prioritized system

$S_i^* = \sum_{m=1}^{i} \lambda_m{}^* c_m$

$\overline{S}_i^* = 1 - S_i^*$

$\wedge_i = \sum_{m=1}^{i} \lambda_m c_m^{(2)}/2$

$M/G/1/P=R$ = priority queue with R priorities

W_q = mean waiting time of a $M/G/1$ queue

$ST_i^1(\underline{\lambda})$ = mean sojourn time of a $M/G/1/P=1$ queue when $\underline{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)$

$ST_i(\underline{\lambda})$ = mean sojourn time of a $M/G/1/P=N$ queue when $\underline{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)$

$TC^1(\underline{\lambda})$ = total delay cost of $M/G/1/P=1$

$TC(\underline{\lambda})$ = total delay cost of $M/G/1/P=N$

$V(\underline{\lambda})$ = total system value

$V_i(\lambda_i)$ = system value attributed to the class-i traffic λ_i

$P_i^1(\underline{\lambda}^*)$ = optimal access charge for a class-i customer under $M/G/1/P=1$

$P_i(\underline{\lambda}^*)$ = optimal access charge for a class-i customer under $M/G/1/P=N$

$\partial TC(\underline{\lambda}^*)/\partial \lambda_i$ = the partial derivative of $TC(\underline{\lambda})$ with respect to λ_i, evaluated at $\underline{\lambda} = \underline{\lambda}^*$

PTD = Priority-and Time-dependent

E_i = expectation operator on service time t of a class-i job

$\prod^i(j)$ = the penalty that a class-i user pays if a class-i user selects class-j priority

TPC_i^1 = the total private cost per class-i job before the transition from non-priority system to the corresponding priority system

$\sigma_j = \sum_{k=1}^{i} \lambda_k c_k^2$