

# Users' Benefit and Provider's Revenue of Content Delivery Services with Priority Control

Kyoko Yamori, Yoshiaki Tanaka  
Kyoko Yamori and Yoshiaki Tanaka:  
Global Information and Telecommunication Institute  
Waseda University  
1-3-10 Nishi-Waseda, Shinjuku-ku, Tokyo, 169-0051  
Yoshiaki Tanaka:  
Japan kyamori@waseda.jp, ytanaka@waseda.jp  
Advanced Research Institute for Science and Engineering  
Waseda University  
17 Kikuicho, Shinjuku-ku, Tokyo, 162-0044 Japan

**Abstract.** Recently, a large amount of content is delivered to users by content delivery services through broad-band networks. The necessity for the service which guarantees quality is increasing in the Internet, because users' requirements are diversified. Differentiated services, which have multiple classes using priority control, are being introduced. The limited network resources can be used effectively by applying priority control in content delivery services. In such services, it is important to predict users' requirements and to provide the services that suit users. In this paper, the user's behaviour in content delivery services with priority control was modeled by using willingness to pay, while the total users' benefit and the provider's revenue in two different content delivery models are maximized. One model is the content delivery service using only streaming, and the other model offers streaming and download services. In this paper, it is proved that optimal pricing for users and that for providers are different.

## 1 Introduction

In recent years, content delivery services are attracting attention as new services with the development of the communication network by ADSL (Asymmetric Digital Subscriber Line) or FTTH (Fibre To The Home). The present IP network is best-

---

*Please use the following format when citing this chapter:*

Yamori, K., Tanaka, Y., 2006, in IFIP International Federation for Information Processing, Volume 226, Project E-Society: Building Bricks, eds. R. Suomi, Cabral, R., Hampe, J. Felix, Heikkilä, A., Järveläinen, J., Koskivaara, E., (Boston: Springer), pp. 240–251.

effort and does not guarantee the quality of service. Since all users are treated equally in best-effort communications, not bandwidth guarantee is offered to particular users. On the other hand, the bandwidth is a key point in real-time applications like video-on-demand, because user's utility is extremely sensitive to quality degradation. Therefore, bandwidth requirements of users have been diversified.

Thus, to meet the diversified bandwidth requirements of users, differentiated services have been examined. The main problems with differentiated services are guaranteed bandwidth and pricing. To study the problems, the relation between user's utility and price for the services has been researched [1-3]. There have also been many studies on quality and utility in fields such as quality evaluation [4,5]. Moreover, there is a considerable amount of research about content delivery with priority control [6,7]. Generally, waiting time and packet loss rate are used for evaluation of content delivery services. However, user's utility does not change linearly against these factors [6]. Therefore, user's utility cannot be evaluated correctly only with these factors alone. We believe there is need for additional evaluation of user' utility in view of the price, application environment, the purpose, etc.

In this paper, the relation between the service quality and the price is quantified using WTP (willingness to pay), which is used to create a model of user's behaviour. We have also considered the pricing so that to maximise total users' benefit and provider's revenue in two different content delivery models with priority control.

## 2 Network Model

In this paper, we consider a network model in which the content is delivered to users through a single transmission line. We assume that content delivery service with priority control is provided on the network. Fig.1 shows the network model.

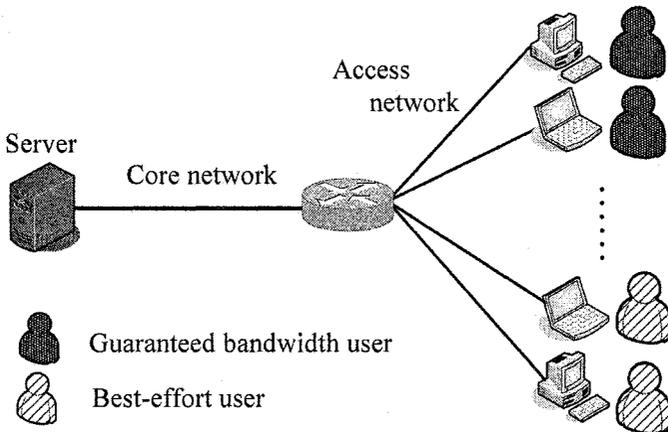


Fig. 1. Network model

Two basic priority classes are provided and stand for high and low user priorities. The high priority class specifies the minimum guaranteed bandwidth, so that even if the network is congested, the minimum bandwidth is guaranteed. When the network utilization is low, the bandwidth is distributed among all users on best-effort basis. When network utilization is high, and the assigned bandwidth to a particular user approaches the minimum guaranteed bandwidth, bandwidth distribution is different and while high priority users get at least the minimum guaranteed bandwidth, low-priority users get only evenly split bandwidth that is left over after high-priority users. Moreover, the maximum bandwidth assigned to the high-priority class is fixed, and, therefore, the number of users in the high-priority class is restricted.

The process of content delivery is shown in Fig.2, along with the pricing table below.

- (1) A delivery request originates from a user and is sent to the content server.
- (2) The information about the present state of priority classes is offered to the user by contents delivery server
- (3) The user selects the delivery class from all offered options.
- (4) Delivery of the content begins.

It is assumed that the price for the delivery of content is directly proportional to the duration of delivery. The basic price is also included for joining in the content delivery service. By paying more money, the user can obtain connection on a higher priority class.

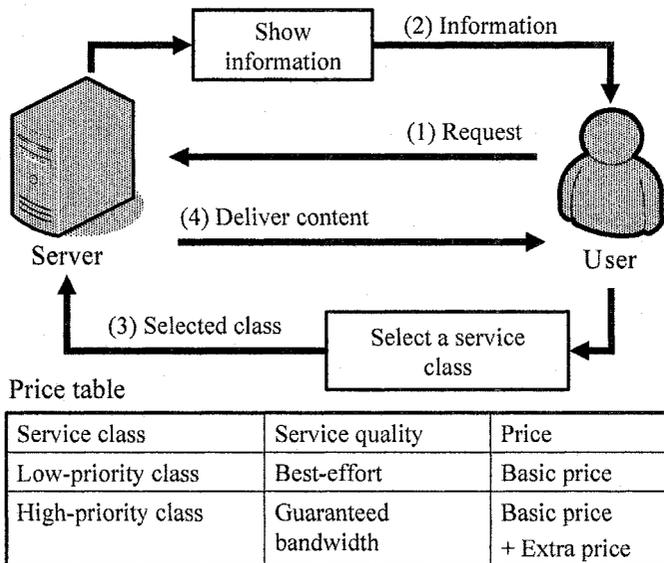


Fig.2. Connection flow and price

### 3 Content Delivery Model

In this paper, two different content delivery models with priority control are defined, that are S-S (streaming vs. streaming) delivery model and S-D (streaming vs. download) delivery model. Table 1 shows details about both models.

**Table 1.** Details about considered content delivery models

S-S delivery model	High-priority class	Streaming service by minimum guaranteed bandwidth.
	Low-priority class	Streaming service by best-effort.
	Information to users from content server	a) Bandwidth for high-priority class. b) Mean bandwidth of low-priority class. c) Extra price for high-priority class.
S-D delivery model	High-priority class	Streaming service by guaranteed minimum bandwidth.
	Low-priority class	Download service by best-effort.
	Information to users from content server	a) Bandwidth for high-priority class. b) Estimation of waiting time. c) Extra price for high-priority class.

In S-S delivery model, since the content is delivered to the low-priority class users using streaming service without the minimum guaranteed bandwidth, when the network is congested, the quality of content deteriorates considerably. On the other hand, in S-D delivery model, the content is delivered to the low-priority class users using download service without guaranteed service, when the networks is congested, the waiting time for complete download becomes excessively long.

To compare S-D and S-S delivery models, the quality of content is guaranteed since the low-priority class is delivered using download service. However, it takes time until download is completed.

We predict the time by using the method of exponential smoothing. Exponential smoothing is a typical time-series-analysis technique which is used to predict future

value from the preceding time series data. The predicted waiting time is defined by Eq.(1),

$$T = \theta T_p + (1 - \theta)T_a, \quad (1)$$

where  $T$  is the predicted waiting time,  $T_p$  is the last predicted waiting time,  $T_a$  is the last actual waiting time. Parameter of  $\theta$  stands for weight, and in this paper is set to  $\theta = 0.5$ .

#### 4 Willingness to Pay

In order to evaluate the differentiated services, one has to match user's behaviour model which determines whether a user joins in service to a price paid for the service. However, since such services do not yet exist in real networks, we need to choose a hypothetical user's behaviour model.

WTP (Willingness to pay) is one of the methods used to define user's behaviour model [8]. WTP is often used for evaluation of services which do not actually exist yet. In the case of using paid services, WTP stands for the maximum amount of money that a user is willing to pay. Therefore, since user's maximum utility can be defined by WTP, it can be used as a measure of user's utility.

Let the price of content be 1. The relation between the WTP and the quality of service is defined by Eqs.(2) and (3),

$$U_{\text{streaming}}(w) = \frac{\alpha_1}{1 + cw^{-\alpha_2}} + 1, \quad (2)$$

$$U_{\text{download}}(t) = \beta_1 \exp[-\beta_2 t] + 1, \quad (3)$$

where  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$ ,  $\beta_2$  are parameters pertaining to a particular user,  $w$  is the available bandwidth of a user,  $t$  is the waiting time, and  $c$  is constant.

Eq.(2) contains the function that defines the relation between WTP and the available bandwidth in streaming service. This function and its parameters are verified by subjectivity evaluation experiment. In the experiment, various quality motion pictures are shown to assessors, which specify how much extra charges can be paid for the guarantee bandwidth [9]. Actually, WTP is influenced by not only bandwidth but jitter and the maximum delay time. To simplify the discussion, we have considered only bandwidth.

Eq.(3) defines the relation between WTP and the waiting time for delivery of 1GByte content using download service. This function and its parameters are tested by a questionnaire.

In the questionnaire, the content size and waiting time are shown to assessors, which are inquired as to how much extra price they are willing to pay considering the waiting time offered to them.

The streaming WTP converges to a certain value, where  $\alpha_1$  is the value to converge to. And  $\alpha_2$  represents sensitivity to the available bandwidth. When  $\alpha_2$  is

large, WTP increases at a higher rate.  $\beta_1$  stands the download WTP in the case when no waiting time occurs.  $\beta_2$  stands for the sensitivity to the waiting time, when  $\beta_2$  is large, WTP decreases at a higher rate. Although the value of WTP itself is depends heavily the particular market, service, and users, it is thought that the tendency is the same.

Figs.3 and 4 show samples of the WTP functions. These samples are obtained with the following parameters:  $\alpha_1=1.3$ ,  $\alpha_2=0.6, 1.0$ ,  $c=3.38$ ,  $\beta_1=1.2$ ,  $\beta_2=0.012, 0.017$ .

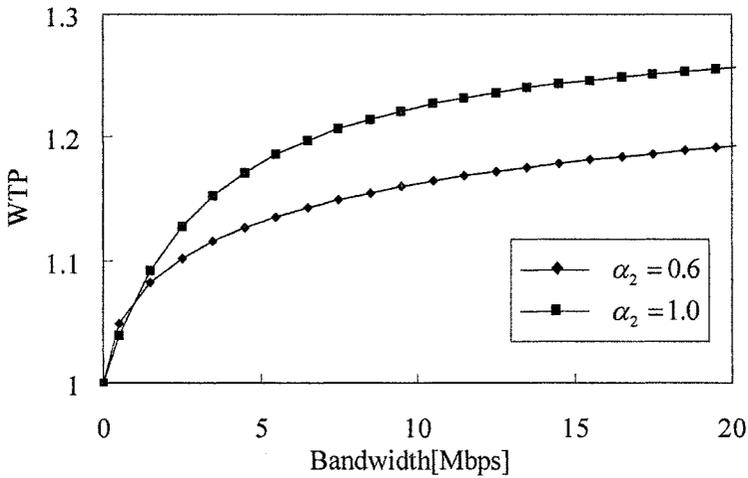


Fig.3. Sample of streaming WTP functions

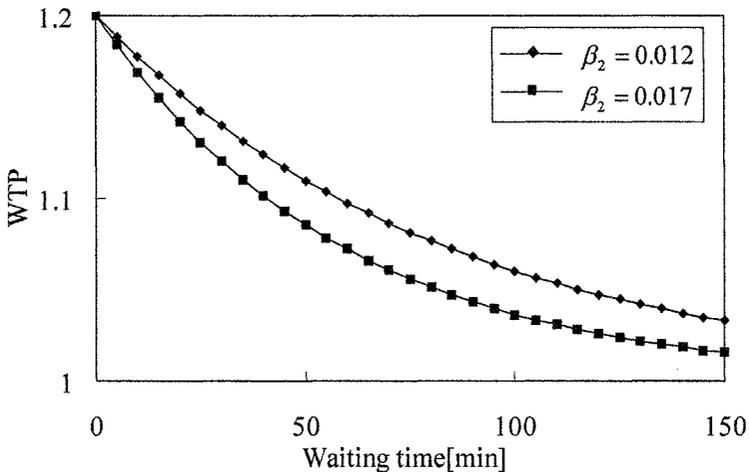


Fig.4. Sample of Download WTP functions

### 5 User's Behaviour Model

In order to define a user's behaviour model, we need to define user's benefit as per Eq.(4),

$$B_i = U_i - p_i \quad (i = 1,2) \quad (4)$$

where  $i$  is the number of the delivery class, i.e.  $i=1$  is the high-priority class and  $i=2$  is the low-priority class.  $U_i$  is WTP of each class and  $p_i$  is a price of each class. The user considers the benefit of each class based on the information received from the content server. After that, the user selects a class which maximises the user's benefit. If  $B_1 \geq B_2$ , the user chooses the high-priority class, if  $B_1 < B_2$ , the user goes for the low-priority class.

$$\begin{cases} B_1 \geq B_2 \Rightarrow (i = 1) \\ B_1 < B_2 \Rightarrow (i = 2) \end{cases} \quad (5)$$

### 6 Simulation Studies

We have evaluated total users' benefit and provider's revenue of both considered delivery models using simulation.

Table 2 shows the conditions of the simulation. The parameters of WTP functions are given to users at random in the range shown in the Table 2. In case of the streaming service, WTP is calculated by the minimum used bandwidth while the content is being delivered. In case of download service, WTP is calculated by the waiting time until the delivery is completed.

Table 2. Simulation setup

Number of content items	1
Transmission time of streaming service	3600 [sec]
Content size	1 [GByte]
Link bandwidth	300 [Mbps]
Minimum guaranteed bandwidth	2 [Mbps]
Maximum bandwidth of the high-priority class	150[Mbps]
Parameters of WTP functions	$0.2 < a_1 < 0.4$
	$0.6 < a_2 < 1.0$
	$c = 3.38$

	$0.18 < \beta_1 < 0.22$
	$0.12 < \beta_2 < 0.17$
Request arrival rate of S-S delivery model	2, 2.5, 3 [times/min]
Request arrival rate of S-D delivery model	2, 2.3, 2.5 [times/min]

Three request arrival rates are given to each delivery models. Total users' benefit per hour and the increase in provider's revenue per hour in the priority control are displayed for each of request arrival rates in Figs. 5-8.

Fig.5 displays the total users' benefit of S-S delivery model, Fig.6 shows the provider's revenue of S-S delivery model, Fig.7 shows the total users' benefit of S-D delivery model, and Fig.8 shows the provider's benefit of S-D delivery model.

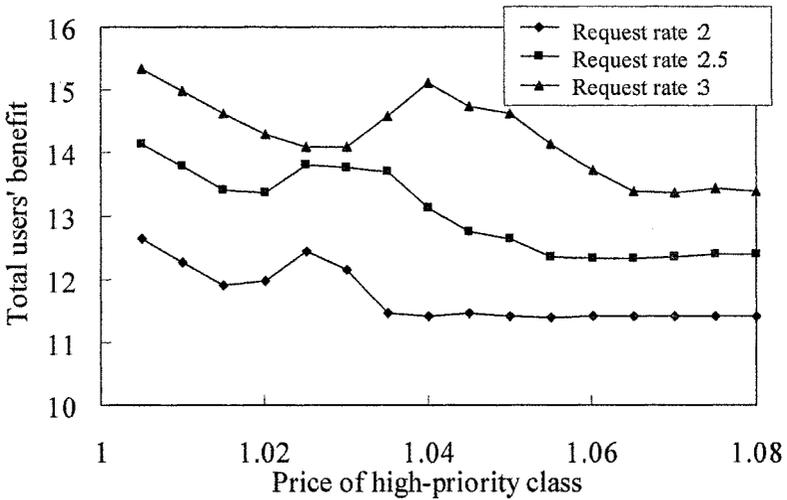


Fig.5. S-S delivery model : Total users' benefit

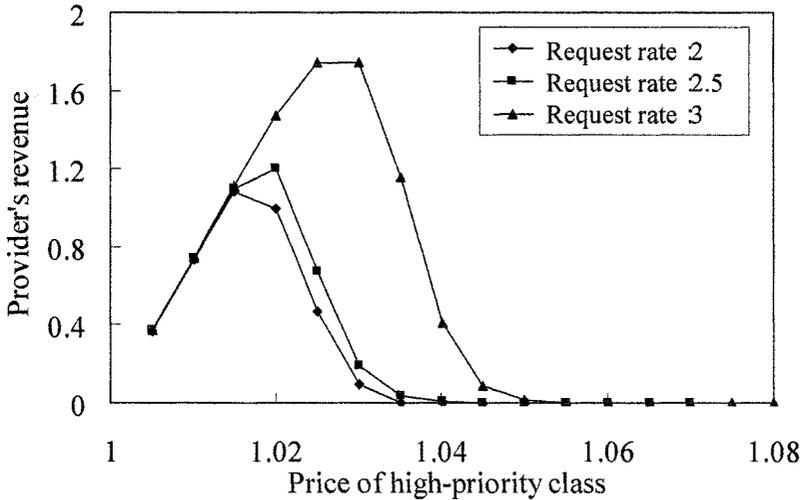


Fig.6. S-S delivery model : Provider's revenue

First, let's consider the results obtained from the S-S delivery model. Figs.5 and 6 exhibit the fact that total users' benefit and provider's revenue change similarly for each request arrival rate. At low prices, the total users' benefit decreases as the price grows. However, when the price is raised even more, total users' benefit is high. This happens because only users who have large WTP join the high-priority class. If the price is raised even more, user's benefit starts to drop. This is because users choose not to join the high-priority class, because the price is too high. As a result, the numbers of the users which join the low-priority class increases. Since the content quality which is delivered to the low-priority class deteriorates, that causes the decrease of total user's benefit as well. Finally, total users' benefit reaches the area where it doesn't change, and that is because users do not join the high-priority class at all. That is, in this area, it is exactly the same as the content delivery without priority control. Considering total users' utility, it can be concluded that it is more beneficial for users to deliver content with priority control than on best-effort basis.

Comparing Fig.5 with Fig 6, provider's revenue reaches the maximum value when the total users' benefit is low. It means that the optimal price for users differs from the optimal price for the provider.

Finally, the results of the S-D delivery model have to be considered. Fig.7 exhibits that the total users' utility increases when the request arrival rate is 2, and decreases when the request arrival rate is 2.3 and 2.5. It means that when the request arrival rate is low, it is better for users to deliver content without priority control, and when the request arrival rate is high, it is better for users to deliver content with priority control. Fig.8 shows that the provider's revenue is considerably high when

the request arrival rate is 2.5. That is, when request arrival rate is high, priority control is very effective for both the users and the provider.

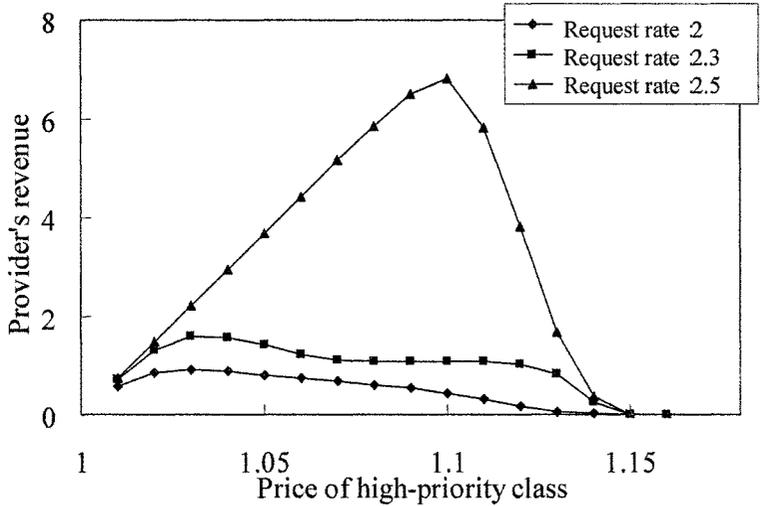


Fig.7. S-D delivery model : Total users' benefit

In the simulation study, content time and size were fixed. As a result of this, the network utilization rate is only affected by the request arrival rate. However, various sizes of content are possible in the actual network. Therefore, it is necessary to examine the relation between network utilization rate, total users' benefit, and provider's revenue.

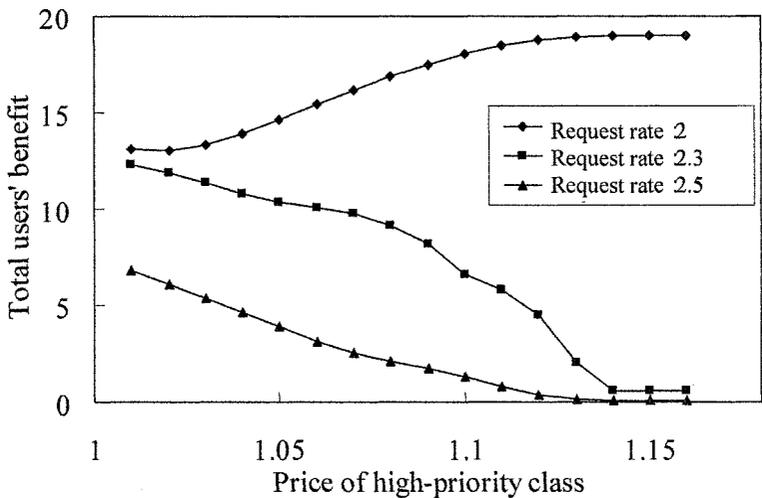


Fig. 8. S-D delivery model : Provider's revenue

## 7 Conclusions

In this paper, we have evaluated total users' benefit and provider's revenue using WTP in content delivery services. As a result, when content is delivered using only streaming service, at any request arrival rate, priority control is very effective for both the users and the provider. On the other hand, when delivery is done over streaming and download services, the effect of using priority control depends on request arrival rate.

As the result of this paper, it is expected that there exists the relation between optimal pricing and traffic load which can be found by adjusting request arrival rate and content size. In order to carry out more flexible optimal pricing, it is necessary to define this relation. Besides, since the optimal price for users differs from the one for the provider, it is necessary to examine the pricing method for each player. These are subjects for the further study.

## Acknowledgment

The authors thank Mr. Shunsuke Ito, who was a graduate student of Waseda University, for his assistance.

## References

1. T. Okamoto and T. Hayashi Analysis of service provider's profit by modeling customer's willingness to pay for IP QoS, Proc. IEEE Globecom 2002, Taiwan, CQRS-04-7, November 2002., pp.1546-1553.
2. I. C. Paschalidis and Y. L. Liu, Pricing in multi-service loss networks: static pricing, asymptotic optimally, and demand substitution effects, IEEE/ACM Trans. Networking, **vol.10**, no.3, June 2002, pp.425-438.
3. I. C. Paschalidis, J. N. Tsitsiklis, Congestion-depend pricing of network service, IEEE/ACM Trans. Networking, **vol.8**, no.2, April 2000., pp.171-184.
4. S. Matsumoto, I. Fukuda, H. Morino, K. Hikichi, K. Sezaki, and Y. Yasuda, The influence of network issues on haptic collaboration in shared virtual environments, Proc. Fifth PHANToM Users Group Workshop, Colorado, the U. S., October 2000.

5. A. Watson and M. Sasse, Measuring perceived quality of speech and video in multimedia conferencing applications, *Proc. ACM Multimedia*, September 1998, pp.55-60, Bristol, United Kingdom.
6. K. Yamori, Y. Tanaka, and H. Akimaru, Pricing method for contents delivery with priority considering request rate, *Proc. 4th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT 2001)*, Kathmandu, Nepal / Atami, Japan, Session 4, November 2001, pp.96-100.
7. N. Yagi, E. Takahashi, K. Yamori, and Y. Tanaka, Efficiency of dynamic pricing in priority-based contents delivery networks, *Proc. 5th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT 2003)*, Noumea, New Caledonia, November 2003, pp.281-286.
8. K. Yamori, H. Ito, and Y. Tanaka, Analysis of willingness to pay in guaranteed multiple-priority service, *Proc. IEEE TENCON 2004 (Region 10 Conference)*, Chiang Mai, Thailand, pp.212-215, November 2004.
9. K. Yamori and Y. Tanaka, Relation between willingness to pay and guaranteed minimum bandwidth in multiple-priority services, *Proc. 10th Asia-Pacific Conference on Communications (APCC 2004)*, Beijing, China, August 2004, pp.113-117.