Pricing and Scheduling in Contents Delivery Networks

Noriyuki YAGI[†], Eiji TAKAHASHI[†], Kyoko YAMORI[†], and Yoshiaki TANAKA[†,††]

[†]Global Information and Telecommunication Institute, Waseda University
1-3-10 Nishi-Waseda, Shinjuku-ku, Tokyo, 169-0051 Japan
Tel: +81 3 5286 3831 Fax: +81 3 5286 3832
E-mail: *nori@suou.waseda.jp, eiji.takahashi@ruri.waseda.jp, yamo@aoni.waseda.jp, ytanaka@waseda.jp*

[††] Advanced Research Institute for Science and Engineering, Waseda University

Abstract: This paper proposes an adaptive pricing system with scheduling to balance the demand for contents and to realize an effective use of resources in contents delivery networks. In the proposed adaptive pricing system, the table of the service levels and prices (tariff) is shown to each user at the start of service and each user chooses one of the service classes. These prices are decided adaptively reflecting the congestion state of the networks. Then, by the proposed scheduling algorithm, these requests are scheduled so as to keep the service level agreements completely.

Key Words: Contents Delivery Networks, Quality of Service, Waiting Time, Pricing, Scheduling

1 Introduction

Contents delivery will be popular in the near future, and its traffic will be dominant in the network. There can be many kind of contents to be delivered in the service, such as, HTML documents, pictures, software, movies and so on.

In contents delivery networks, users often make some requests for large-size contents to be delivered. Traffic of these large-size contents causes a heavy congestion especially at peak usage hours. Therefore, there is a necessity for carriers to accommodate extra capacity to meet peak usage demands; even if many resources are not used in off-peak hours. To avoid building additional capacity to meet peak usage demands, price incentives are provided as an effort to shift some of the demands from peak to nonpeak periods in an efficient manner [1].

This paper proposes an adaptive pricing system with scheduling to balance the demand for contents and to realize an effective use of resources in contents delivery networks. In the proposed adaptive pricing system, the table of service levels and prices (tariff) is shown to each user at the start of each service and each user chooses one of the service classes. These prices are decided adaptively reflecting the congestion state of the networks. Then, by the proposed scheduling algorithm, these requests are scheduled so as to keep the service level agreements completely.

2 Market-based Scheduling System in Contents Delivery Networks

2.1 Framework

Market-based scheduling system that is shown in Figure 1 is considered in this paper. First, a request of each user is sent to the server. Then, the server set up a tariff according to the congestion state of the networks and shows the tariff to each user. Each user chooses one of the service classes shown in the tariff. These requests are scheduled so as to keep the service level agreements (waiting time, in this paper) shown in the tariff. In the tariff, for example, a price of the class where contents will be transmitted right now, after 1 hour or after 2 hour is shown. Hereafter, we consider the case of two service classes to make the problem simple.

By setting prices adaptively reflecting the congestion state of the networks, requests are admitted based on the current valuation (willingness to pay) for contents delivery services. The charge is close to the low price end when the networks are not congested. On the other hand, the charge is close to the high price end when the networks are congested.

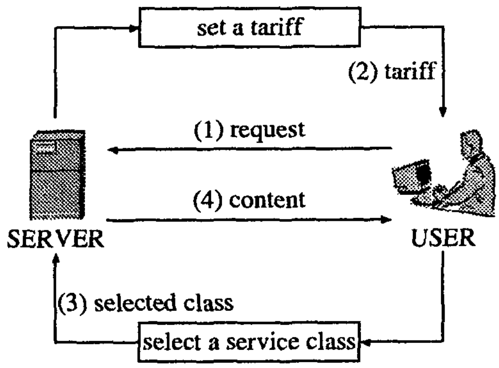

Figure 1: System model.

2.2 Spot Pricing Without Scheduling

First, the spot pricing without scheduling is discussed. This pricing method is often used successfully in telephone services by telecommunication common carriers. The network resources will be used more efficiently compared with the fixed pricing.

Let $i = \{1, 2\}$ and $C_{0,i}$ denote the ID of each service class and the minimum price, the price when the network is not congested. Let t denote the arrival time of the request. Then, the price C_i is decided as:

$$C_i = \begin{cases} C_{0,i} & 0 < t \leq 18, \\ C_{0,i} + a_i(t-18) & 18 < t \leq 21, \\ C_{0,i} - a_i(t-24) & 21 < t \leq 24, \end{cases} \quad (1)$$

where a_i is a constant value set up for each service class individually. Example of the tariff is shown in Table 1, where W_i denotes the waiting time. In this method, the price is predetermined. Then it is impossible to deal with traffic variation. Therefore, sometimes, it is difficult to guarantee the service level agreement shown at the starting point of the service.

Table 1: Tariff.

Service Class i	Price	Waiting Time
1	C_1	W_1
2	C_2	W_2

2.3 Dynamic Pricing with Scheduling

Secondly, the dynamic pricing with scheduling is discussed. In this method, the price is set up reflecting the congestion state of the network. By this pricing method, the service level agreement can be guaranteed. Transmission of contents is controlled according to the delivery schedule shown in Table 2, where $j = \{1, 2, \cdots, N\}$ and S_j denote the number of each requests and the data size of each content. The delivery schedule includes data of requests ID, size of each content and delivery time requested.

These requests are sorted by decreasing the delivery time $t + W_i$ at the any time a request occurs, where t and W_i denote a current time and the waiting time shown in the tariff. In the case of 2 service classes, a requested data of Service Class 1 preempts into the queue according to $t_l < t + W_1 < t_{l+1}$, where $l \in j$. Then the queue length of Service Class 1 data is given as:

$$Q_1 = \sum_{i=1}^{l} S_j / B, \quad (2)$$

where B denotes the bandwidth of a bottleneck link in the contents delivery network. On the other hand, any data of Service Class 2 request will be added to the end of the queue, the queue length of Service Class 2 requests is given as:

$$Q_2 = \sum_{i=1}^{n} S_j / B. \quad (3)$$

Prices are set by substituting the queue length of each service given by Eqs. (2) and (3) for the function shown in Figure 2.

Table 2: Delivery schedule.

Request No. j	Data Size (MB)	Delivery Time
1	S_1	t_1
2	S_2	t_2
...	...	...
N	S_N	t_N

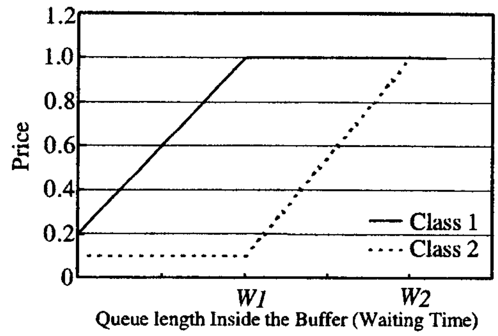

Figure 2: Price setting function.

3 Simulation Models

3.1 Network Model

To make the problem simple, a single link that has a total capacity of B as shown in Figure 3 is considered in the simulation.

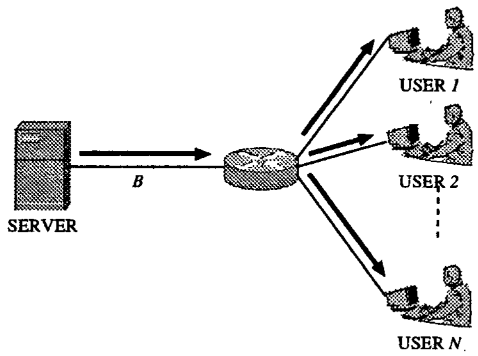

Figure 3: Network model.

3.2 Traffic Model

To simplify the calculation, it is assumed that arriving packets follow a Poisson distribution. Let us assume that the arrival rate of requests for contents is given as shown in Figure 4, where λ_0 is arrival rate during the period from 0:00 to 18:00 [2]. The pattern of call arrival rate shown in Figure 4 was given by two test services of VoD [3], [4]. Then, λ_0 is decided as:

$$\lambda_0[1/\text{sec}] = \text{R}[1/\text{day}]/(36 \times 3600), \quad (4)$$

where R is number of average demands with one day.

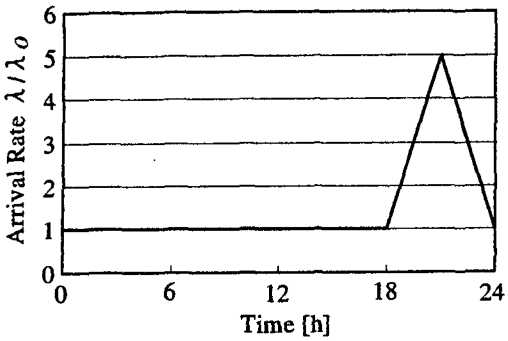

Figure 4: Distribution of standardized mean arrival rate.

3.3 User's Utility

Utility functions are widely used in pricing theory; some common assumptions about the function include concavity and strict monotonicity. Let W and T denote the waiting time shown in the tariff and the time until the user need the content, respectively.

Let us define the user's utility function as [5]:

$$U(W) = \begin{cases} DW^{-k}, & W > T \\ 1, & 0 \leq W \leq T. \end{cases} \quad (5)$$

The parameter D is a scale factor which satisfies $U(T) = DT^{-k} = 1$ in this paper. The parameter k is expressing the sensitivity of each user against the waiting time. When the value of k is larger, the user's utility will be decreased rapidly when the delivering time is extended over the time required. On the other hand, if the value of k is close to 0, a user's utility will hardly be influenced by waiting time. At the time of $k = 0$, a user's utility is not influenced at all by waiting time. The parameter $k > 0$ may be statistically estimated by opinion tests ($k = 0.5$ in this paper) [6].

3.4 User's Behaviour

Let us suppose each user behaves as followings in corresponding to prices. Let us define a payoff H_s to user s as

$$H_s = U(W_i) - C_i, \quad (6)$$

where W_i and C_i denote the waiting time and asking price of Service Class i which were shown in the tariff. Each user maximizes the payoff to him as shown in Eqn. (7) when he determines to choose the service class [7].

$$\max_i U(W_i) - C_i, \quad (7)$$

where $H_s > 0$. If $H_s \leq 0$, user s will cancel his request.

3.5 Parameters

Individual parameters used in the simulation are shown in Table 3.

Table 3: Parameters used in the simulation.

Parameter	Value
k	0.5
T	random
R	800
number of service classes	2
waiting time of Service Class 1	1 hour
waiting time of Service Class 2	2 hours
size of each contents	650 MB
capacity of bottleneck link	50 Mbps

4 Simulation Studies

4.1 Effectiveness of Spot Pricing Without Scheduling

Initial tariff settings in the case of spot pricing are shown in Table 4. The parameter a_1 is set for $C_1 = 0.9$ when $t = 21$ in the case of Service Class 1. The parameter a_2 is set for $C_2 = 0.5$ when $t = 21$ in the case of Service Class 2. Simulation results are shown in Figure 5.

Figure 5 shows the relationship between the waiting time and a time. Comparing with the case of a fixed price service, we can reduce the waiting time effectively. In the case of the fixed pricing, the price is fixed not regarding on the changes of request rates. This is because we can balance the demands by using spot pricing system. By setting the price adaptively, the demands were shifted from peak to nonpeak periods or some requests are canceled in an efficient manner. However, the spot pricing service is not satisfying the service level agreements shown in the tariff at the start of the service to each user. To address this, we proposed a dynamic pricing system with scheduling, where all requests are scheduled so as to keep the service level agreements completely.

Table 4: Initial tariff settings in the case of spot pricing without scheduling.

Service Class	Price	Waiting Time (hour)
1	0.5	1
2	0.3	2

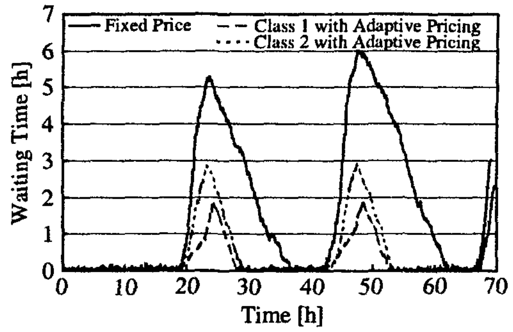

Figure 5: Waiting time in the case of spot pricing without scheduling.

4.2 Effectiveness of Dynamic Pricing with Scheduling

Simulation results are shown in Figure 6.

Figure 6 shows the relationship between the waiting time and a time. As all requests are scheduled in the system, the service level agreements are kept completely.

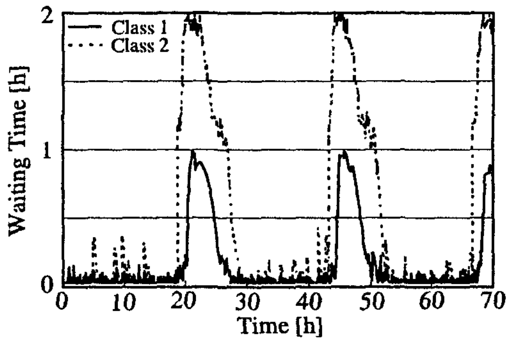

Figure 6: Waiting time in the case of dynamic pricing with scheduling.

5 Conclusion

This paper proposed an adaptive pricing system to balance the demand for contents. First, we compared the case of a fixed pricing and the case of spot pricing without scheduling. By using spot pricing system, the demands were shifted from peak to nonpeak periods in an efficient manner. As a result, waiting time can be reduced drastically. However, the spot pricing system sometimes does not keep the service level agreements. Therefore, we proposed dynamic pricing with scheduling system. In this system, prices were set up based on the queue length in a buffer. The service level agreements are kept completely by the proposed system.

In this examination, the easy model was used to simplify the calculation. It is left for further study to use the more realistic model. We have to take account of background traffic, however, it is difficult to keep the service level agreements completely or we have to predict the background traffic exactly. The assumptions of users behaviours in corresponding to prices should be verified statistically. This point should be clarified by the opinion test, etc.

Acknowledgements

This study was supported by the Waseda University Grant for Special Research Projects (Individual Research: 2002A-600).

References

[1] J. K. MacKie-Mason and H. R. Varian, "Pricing the Internet," Public Access to the Internet, The MIT Press, 1995.

[2] N. Kamiyama, "An efficient transmission protocol for multicast video-on-demand system(in Japanese)," IEICE Technical Report, SSE2000-252, IN2000-208, March 2001.

[3] Haar PG de, et al., "DIAMOND Project: Video-on-demand system, and trials," Eur. Trans. Telecommun., vol.8, no.4, pp.337-344, 1997.

[4] Bell Atlantic, "Fact sheet: Results of Bell Atlantic video services video-on-demand market trial," Trial Results, 1996.

[5] K. Yamori, Y. Tanaka, and H. Akimaru, "Price optimization of contents delivery systems with priority," IEEE International Conference on Networking (ICN 2001), Colmar, France, Lecture Notes in Computer Science 2093, Springer, pp.65-74, July 2001.

[6] K. Nomura, K. Yamori, E. Takahashi, T. Miyoshi and Y. Tanaka, "Waiting time versus utility to download images," 2001 Asia Pacific Symposium on Information and Telecommunication Technologies (APSITT2001), Kathmandu, Nepal/ Atami, Japan, pp. 128-132, November 2001.

[7] R. Gibbons, "Game theory for applied economists (in Japanese)," Sobunsha, July 1995.