Revenue-aware Resource Allocation Scheme in Multiservice IP Networks

Jian Zhang, Timo Hämäläinen, Jyrki Joutsensalo
Dept. of Mathematical Information Technology
University of Jyväskylä, FIN-40014 Jyväskylä, Finland

***Abstract*— In the future IP networks, a wide range of different service classes must be supported and different classes of customers will pay different prices for their used network resources based on Service-Level-Agreements. In this paper, we link resource allocation scheme with pricing strategies and explore the problem of maximizing the revenue of network providers by resource allocation among multiple service classes under a certain Service-Level-Agreement and a given amount of network resources. A revenue-aware resource allocation scheme is proposed under linear pricing strategy, which has the closed-form solution to the optimal resource allocation for maximizing the revenue per time unit gained in a network node. The optimal allocation scheme is derived from revenue target function by Lagrangian optimization approach.**

Keywords: Network resource allocation, QoS, Pricing strategy, Revenue maximization.

I. Introduction

Resource allocation in the multiservice communication networks presents a very important problem in the design of the future multi-class Internet. The main motivation for the research in this field lies in the necessity for structural changes in the way the Internet is designed. The current Internet offers a single class of 'best-effort' service, although some traffic prioritization will be active in the new network router implementations. The future IP networks must carry a wide range of different traffic types being still able to provide performance guarantees to real-time sessions such as Voice over IP (VoIP), Video-on-Demand (VoD), or Video-Conferencing. Efficient and effective communication needs careful Quality of Service (QoS) design by means of appropriate resource allocation among competing traffic flows with different service classes. On the other hand, for the future multi-class Internet, users will have to pay the network providers based on pricing strategies agreed in their Service-Level-Agreements. Obviously, the pricing strategy will specify the relationship between the price paid by each class of users and the QoS (e.g., delay, jitter) provided by the network provider, which normally states that the network provider will get a revenue when the offered QoS meets the minimal performance requirement and suffer a penalty when the offered QoS fails to meet that. Network designers are facing a complicated problem of optimizing the network control to satisfy both the issue of performance guarantees for multi-class traffics and the issue of maximizing the revenues of service providers.

Pricing research in the networks has been quite intensive during the last few years (e.g., [6], [3], [10], [11]). Also a lot of work (e.g., [14], [9], [7]) has been done concerning the issues of resource allocation and fairness in a single-service environment. But the combination of pricing strategies and resource allocation among multiple service classes have not been analyzed widely. A number of works [12], [2], [13], [8] recently use end-users' utility as the maximizing objective for resource allocation schemes. All of these approaches have a common objective of maximizing the network performance in terms of the users' utility. Our research differs from these studies by linking resource allocation scheme with pricing strategies of multiple service classes to maximize the revenue gained under a certain amount of resources. A revenue-maximizing pricing scheme for the service provider is presented in [1], where a noncooperative (Nash) flow control game is played by the users (followers) in a Stackelberg game with the goal of setting a price to maximize revenue. Our scheme proposed in this paper is to maximize the revenue under given pricing strategies by the optimal resource allocation among competing traffic flows from different service classes.

This paper extends our previous QoS and pricing research ([5], [4]) and takes into account revenue maximization issue by introducing new revenue-aware resource allocation scheme into multiservice IP networks. In a network node supporting multiple service classes, packets are queued in a multi-queue system, where each queue corresponds to one service class. Based on a pricing strategy which specifies the relationship between the paid prices and the offered service performances, the network provider will get a revenue or suffer a penalty whenever serving one incoming packet. In this paper, a revenue-aware resource allocation scheme is proposed under linear pricing strategy, which has the closed-form solution to the optimal resource allocation derived from revenue target function by Lagrangian optimization approach.

The rest of the paper is organized as follows. In Section 2, three pricing strategies (linear, flat and piecewise linear) are presented and the linear one is generally defined. Revenue-aware resource allocation scheme is derived in Section 3, where the optimal allocating solution is given for maximizing the revenue of service providers. Section 4 contains simulation part demonstrating the revenue-maximizing ability of our proposed resource allocation scheme. Finally, in Session 5,

The work of Jian Zhang reported in this paper was supported partly by a scholarship grant from Nokia Foundation

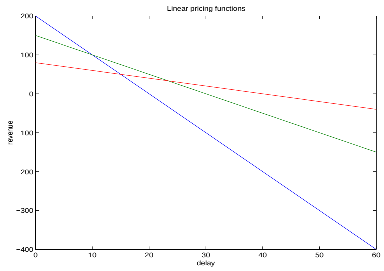

Fig. 1. Three linear pricing functions. Horizontal axis: delay; vertical axis: price.

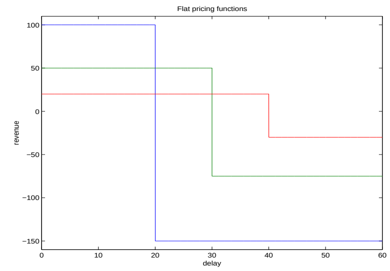

Fig. 2. Three flat pricing functions. Horizontal axis: delay; vertical axis: price.

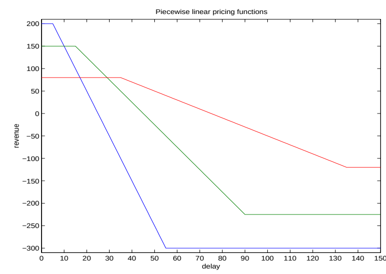

Fig. 3. Three piecewise linear pricing functions. Horizontal axis: delay; vertical axis: price.

we present concluding remarks.

II. Pricing Strategy

Three pricing strategies are presented here which are believed to be the most used ones. First some parameters and notions are defined. We consider a network node which supports multiple service classes. There incoming packets are queued in a multi-queue system (each queue corresponds to one service class) and the resources in the network node (e.g. processor capacity and bandwidth) are shared among those service classes. The number of classes is denoted by *m*. Literature usually refers to the gold, silver and bronze classes; in this case, $m = 3$. The metric of service performance considered in this paper is packet delay which is most concerned by end users. The packet delay of class *i* in the network node is referred to as $d_i(t)$. Hereafter, time index *t* is dropped for convenience. For each service class, a pricing function $r_i(d_i)$ is defined to rule the relationship between the QoS (packet delay here) provided by a network provider and the price paid by its customers. Obviously, it is non-increasing with respect to the delay d_i. Some examples of pricing functions are given in Figs. 1, 2, and 3, which show the most used pricing strategies: linear, flat and piecewise linear functions, respectively. In this paper, our study concentrates on the revenue-maximizing issue under linear pricing functions and the analysis under flat pricing strategy is postponed to its sequel. The solution to the piecewise linear pricing strategy is a straightforward extension to the above two cases. Specifically, Linear pricing strategy for class *i* is characterized by the following definition.

Definition 1: *The function*

$$r_i(d_i) = b_i - k_i d_i, i = 1, 2, ..., m, b_i > 0, k_i > 0 \quad (1)$$

is called *linear pricing function*, where b_i and k_i are positive constants and normally $b_i \geq b_j$ and $k_i \geq k_j$ hold to ensure differentiated pricing if class *i* has higher priority than class *j* (in this paper, we assume that class 1 is the highest priority and class *m* is the lowest one).

Fig. 1 depicts three linear pricing functions for gold, silver and bronze classes and it is commented in detail below. For gold class, the pricing function $r_1(d_1) = 200 - 10d_1$ means that when the delay d_1 is smaller, the price paid by the gold class of customers is higher - in this case, maximally 200 units of money. It is natural that for the highest priority class, constant shift b_1 is selected to be the highest which determines the maximum price paid by customers. On the other hand, the penalty paid to the customers with the highest priority class is also the highest if the network provider fails to meet the minimum delay requirement (in this case, 20 time units for the gold class); the growing rate of penalty along with the delay depends on the slope k_1 and it is also the highest. Same observations hold for silver and bronze classes. For bronze class, $r_3(d_3) = 80 - 2d_3$ means that the price paid by that class of customers is maximally 80 units of money, i.e., constant shift b_3 is the lowest. On the other hand, the penalty paid to the bronze class of customers is the lowest when failing to meet its minimal delay requirement (in this case, 40 time units for the bronze class) and the growing rate of penalty is also the lowest. For example, assume that a packet experiences a delay of 30 time units in the network node. If the packet belongs to the gold class, i.e., $d_1 = 30$, then $r_1(d_1) = r_1(30) = 200 - 10 * 30 = -100$ means the network provider should pay 100 units of money to the customer as the penalty; whereas, if the packet belongs to the bronze class, i.e., $d_3 = 30$, then $r_3(d_3) = r_3(30) = 80 - 2 * 30 = 20$ means the network provider will get 20 units of money for serving that packet. These are actually what we expect based on the requirement of Service-Level-Agreement.

III. Revenue-aware Resource Allocation Scheme

Let us consider a network node which has the resource of *C* bit/s (processing capacity and/or bandwidth) and will support *m* service classes totally. The traffic flows fed into the network node are Poisson streams with arrival rate λ_1, λ_2, ..., λ_m, respectively. We assume that the distribution of packet length of all classes is exponential and use $\bar{L}_i$ (bits) to denote the mean packet length of class *i*. The portion of the resource allocated to class *i* is denoted as $w_iC, i = 1, 2, ..., m$ and w_i is referred to as the weight allotted to class *i*. Without loss of generality, only non-empty queues are considered, and thus $w_i \neq 0$. If some weight $w_i = 1$, then $m = 1$. Therefore, the natural constraint for the weights is $\sum_{i=1}^{m} w_i = 1, w_i \in (0, 1]$.

When the weight assigned to class *i* is w_i, class *i* can be guaranteed to have a portion of the resource w_iC; and the packets of class *i* arrive at queue *i* with rate λ_i. Therefore, the mean packet delay of class *i* $\bar{d}_i$ in the network node can be denoted as

$$\bar{d}_i = \frac{1}{\frac{w_iC}{\bar{L}_i} - \lambda_i} = \frac{\bar{L}_i}{w_iC - \lambda_i\bar{L}_i} \tag{2}$$

based on queueing theory. The natural constraint of Eq. (2) is $w_iC > \lambda_i\bar{L}_i$ due to the fact that delay can not be negative.

The metric of revenue used in this paper is the revenue gained per time unit since a network provider will obtain a revenue or penalty whenever one packet is served. Unless stated otherwise, we shall hereafter refer to the revenue per time unit as revenue. We use the mean packet delay $\bar{d}_i$ in Eq. (2) to estimate the real packet delay d_i. Then the revenue gained in a network node *F* may be defined as follows when the linear pricing function in Eq. (1) is deployed:

$$F = \sum_{i=1}^{m} \lambda_i r_i(d_i) = \sum_{i=1}^{m} \lambda_i (b_i - \frac{k_i\bar{L}_i}{w_iC - \lambda_i\bar{L}_i}) \tag{3}$$

As a result of the above definition, the issue of revenue maximization under linear pricing strategy by resource allocation can be formulated as follows:

$$max \qquad F = \sum_{i=1}^{m} \lambda_i (b_i - \frac{k_i\bar{L}_i}{w_iC - \lambda_i\bar{L}_i}) \tag{4}$$

$$s.t. \qquad \sum_{i=1}^{m} w_i = 1, 0 < w_i \leq 1 \tag{5}$$

$$w_iC > \lambda_i\bar{L}_i \tag{6}$$

Theorem 1. *For linear pricing strategy, the globally maximum revenue F gained in a network node is achieved by using the following optimal resource allocation scheme*

$$w_i = \frac{\sqrt{\lambda_i k_i \bar{L}_i}(C + \frac{\sum_{j=1}^{m}\sqrt{\lambda_j k_j \bar{L}_j}}{\sqrt{\lambda_i k_i \bar{L}_i}}\lambda_i\bar{L}_i - \sum_{j=1}^{m}\lambda_j\bar{L}_j)}{C\sum_{j=1}^{m}\sqrt{\lambda_j k_j \bar{L}_j}} \tag{7}$$

for i = 1, 2, ..., m and it is unique when $w_i \in (0, 1]$.

Proof: Based on Equations (4) and (5), we can construct the following Lagrangian equation.

$$P = \sum_{i=1}^{m} \lambda_i (b_i - \frac{k_i\bar{L}_i}{w_iC - \lambda_i\bar{L}_i}) + \sigma(1 - \sum_{i=1}^{m} w_i) \tag{8}$$

Set partial derivatives of *P* in Eq. (8) to zero:

$$\frac{\partial P}{\partial w_i} = \frac{\lambda_i k_i \bar{L}_i C}{(w_iC - \lambda_i\bar{L}_i)^2} - \sigma = 0. \tag{9}$$

It follows that

$$\sigma = \frac{\lambda_i k_i \bar{L}_i C}{(w_iC - \lambda_i\bar{L}_i)^2} \tag{10}$$

leading to the solution

$$w_i = \sqrt{\frac{\lambda_i k_i \bar{L}_i}{C\sigma}} + \frac{\lambda_i\bar{L}_i}{C}, i = 1, 2, ..., m. \tag{11}$$

Substituting Eq. (11) to Eq. (5), we get

$$\sqrt{\sigma} = \frac{\sum_{i=1}^{m}\sqrt{\lambda_i k_i \bar{L}_i C}}{C - \sum_{i=1}^{m}\lambda_i\bar{L}_i} \tag{12}$$

And when $\sqrt{\sigma}$ in Eq. (12) is substituted to Eq. (11), the closed-form solution in Eq. (7) is obtained.

Because of the constraint in Eq. (6) $w_iC > \lambda_i\bar{L}_i$, obviously,

$$\sum_{j=1}^{m} w_jC = C > \sum_{j=1}^{m}\lambda_j\bar{L}_j \tag{13}$$

Hence, the closed-form solution in Eq. (7) $w_i > 0$. Moreover, based on (13), the following inequality holds

$$\lambda_i\bar{L}_i - \frac{\sqrt{\lambda_i k_i \bar{L}_i}\sum_{\substack{j\neq i\\ j=1}}^{m}\lambda_j\bar{L}_j}{\sum_{\substack{j\neq i\\ j=1}}^{m}\sqrt{\lambda_j k_j \bar{L}_j}} \leq C$$

leading to in Eq. (7) the numerator less than the denominator. Hence, we can conclude that $0 < w_i \leq 1$.

To prove that the closed-form solution in Eq. (7) is the only and optimal one in the interval (0, 1], we consider second order derivative of *P*.

$$\frac{\partial^2 P}{\partial w_i^2} = -\frac{2\lambda_i k_i \bar{L}_i c^2}{(w_iC - \lambda_i\bar{L}_i)^3} < 0 \tag{14}$$

due to the constraint $w_iC > \lambda_i\bar{L}_i$ in (6). Therefore, the revenue per time unit *F* is strictly convex with the allotted set of weights $\{w_1, ..., w_i, ..., w_m\}$ in the interval $0 < w_i \leq 1$, having one and only one maximum. This completes the proof. **Q.E.D.**

In addition, the theoretical maximum revenue gained by a network provider can be calculated as follows.

Theorem 2. *When the optimal resource allocation scheme is deployed according to Theorem 1, the theoretical maximum revenue obtained in a network node is*

$$F_{max} = \sum_{i=1}^{m}(\lambda_i b_i) - \frac{(\sum_{i=1}^{m}\sqrt{\lambda_i k_i \bar{L}_i})^2}{C - \sum_{i=1}^{m}\lambda_i\bar{L}_i} \tag{15}$$

Proof: When the optimal weights in Eq. (7) are substituted to Eq. (3), the theoretical maximum of *F* is

$$\begin{aligned} F_{max} &= \sum_{i=1}^{m}(\lambda_i b_i - \frac{\lambda_i k_i \bar{L}_i \sum_{i=1}^{m}\sqrt{\lambda_i k_i \bar{L}_i}}{\sqrt{\lambda_i k_i \bar{L}_i}(C - \sum_{i=1}^{m}\lambda_i\bar{L}_i)}) \\ &= \sum_{i=1}^{m}(\lambda_i b_i - \frac{\sqrt{\lambda_i k_i \bar{L}_i}\sum_{i=1}^{m}\sqrt{\lambda_i k_i \bar{L}_i}}{C - \sum_{i=1}^{m}\lambda_i\bar{L}_i}) \\ &= \sum_{i=1}^{m}(\lambda_i b_i) - \frac{(\sum_{i=1}^{m}\sqrt{\lambda_i k_i \bar{L}_i})^2}{C - \sum_{i=1}^{m}\lambda_i\bar{L}_i} \end{aligned}$$

Q.E.D.

IV. Simulations

In this section we present the simulation results which demonstrate the effectiveness of our resource allocation scheme for maximizing the revenues of network providers under linear pricing strategy and a given amount of resources. A number of simulations have been conducted under different parameter settings. In each case, we first numerically determine the optimal allocation scheme using Theorem 1, and then we investigate through simulations the benefits of the optimal scheme by comparing the revenues obtained under the optimal allocation with those obtained under a natural scheme of proportional allocation as well as the theoretical maximum revenues. A representative set of these simulations are presented herein. Throughout this section, we shall focus on a network node where its resource C equals 10^6 bit/s and the number of service classes supported $m = 3$ (namely, gold, silver and bronze classes). The base arrival rates and the mean packet lengths of the above three classes are provided in Table 1. A multiplicative *load factor* $\rho > 0$ is used to scale these base arrival rates to consider different traffic intensities; i.e., $\lambda_j \rho$ will be used in the simulations as class-j arrival rate. As mentioned above, we use a scheme that proportionally allocates the resource among all service classes for comparison with our revenue-aware resource allocation scheme. Specifically, the *proportional* scheme allots the weight of class i as follows: $w_i = \frac{\lambda_i \bar{L}_i}{\sum_{j=1}^{m}(\lambda_j \bar{L}_j)}, i = 1, 2, ..., m$. Note that this proportional scheme is a natural way to allocate network resources.

A. The first set of simulations

In the first set of simulations, the parameters related to three used linear pricing functions are summarized as follows: $b_1 = 200$, $k_1 = 10000$, for gold class, $b_2 = 150$, $k_2 = 5000$, for silver class, and $b_3 = 80$, $k_3 = 2000$ for bronze class (note that the time unit is second hereafter).

First we investigate the evolution of revenue along with the time under our optimal allocation scheme and the proportional scheme. In this case, the base arrival rates in Table 1 are used and one allocation scheme with a set of given weights ($w_1 = 0.60$, $w_2 = 0.25$, $w_3 = 0.15$, referred to as given scheme) is also used for comparison with our scheme. Fig. 4 presents the simulation results, where the x-axis represents the time (the measurement period is 100 seconds here) and the y-axis represents the revenue per second. It is observed that the largest revenue is achieved under our optimal allocation scheme compared with those achieved under the proportional scheme and the given scheme. Moreover, the simulated revenue under our optimal scheme is quite close to the theoretical maximum revenue calculated by Eq. (15). Since the parameters used in Eq. (15) are constant in this case, the theoretical maximum remains unchanged; whereas, as the real packet delay is variable, the simulated revenue varies along with the time. Fig. 4 shows that the revenue obtained under our optimal allocation scheme is very close to the theoretical maximum, which demonstrates the effectiveness of our scheme for revenue maximization. Additionally, the revenue obtained under the proportional scheme is larger than

	$i = 1$ (gold class)	$i = 2$ (silver class)	$i = 3$ (bronze class)
λ_i (packets/s)	10	15	20
L_i (bits)	3360	3360	3360

TABLE I

The base parameters for packet traffic

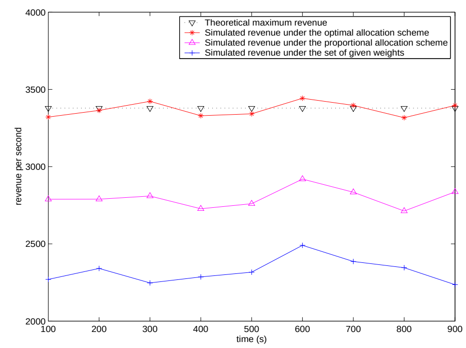

Fig. 4. Revenue comparison as function of time, for the case, load factor $\rho = 1$ and $b_1 = 200$, k_1=10000, $b_2 = 150$, $k_2 = 5000$, $b_3 = 80$, $k_3 = 2000$.

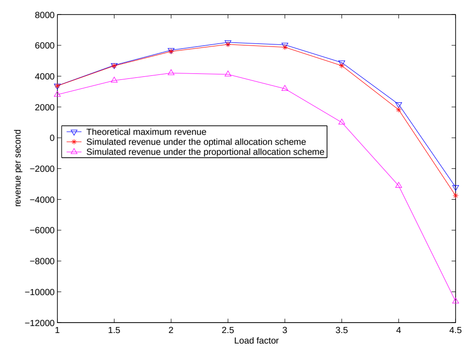

Fig. 5. Revenue comparison as function of load factor ρ, for the case, $b_1 = 200$, k_1=10000, $b_2 = 150$, $k_2 = 5000$, $b_3 = 80$, $k_3 = 2000$.

the one obtained under the set of given weights, which shows that the proportional allocation scheme is somehow acceptable one in this case.

Next we examine the performance of our optimal allocation scheme for the case that the same pricing functions are used and different traffic intensities are fed into the network node. Fig. 5 shows the simulation results, where the x-axis represents the load factor and the y-axis represents the revenue.

We can see in Fig. 5 that the revenues obtained under our optimal allocation scheme are extremely close to those theoretical maximums under light and medium loads, and both are growing almost linearly. This is as expected because few penalties will be incurred under such loads. Under heavy loads both curves start to level off as the penalties start to grow faster than the revenues. Compared with our optimal scheme, the proportional allocation scheme achieves less revenues under all traffic loads. Although the revenue curve of the proportional scheme also grows under light loads, it starts to decrease much earlier as the penalties incurred under the proportional scheme are much larger than the ones under our optimal scheme when the same workload is fed into the network node.

B. The second set of simulations

In the second set of simulations, the same simulations are made under three different linear pricing functions: $b_1 = 200$, $k_1 = 5000$, for gold class, $b_2 = 120$, $k_2 = 2000$, for silver class, and $b_3 = 40$, $k_3 = 500$, for bronze class, to evaluate the performance robustness of our optimal scheme for revenue maximization. Figs. 6 and 7 present the simulation results.

It is observed in Fig. 6 that the revenue obtained under our optimal scheme is the largest and it is also close to the theoretical maximum by Eq. (15); whereas, the revenue obtained under the proportional scheme is less than the one obtained under the set of given weights in this case. Since the slope k_i of class *i* in this case is less than the one used in the first set of simulations, the revenue of class *i* will decrease more slowly along with the increase of delay in this case, leading to the revenue curves in Fig. 7 still grows under heavier loads compared with the ones in Fig. 5. The point is the largest revenue is obtained by our optimal scheme under all traffic loads and it is also very close to the curve of theoretical maximum revenue. Therefore, the robustness of the revenue-maximizing ability of our optimal allocation scheme is demonstrated under linear pricing strategy.

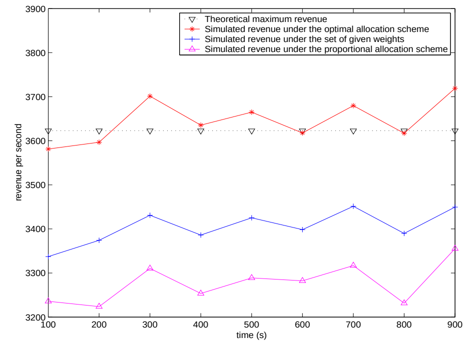

Fig. 6. Revenue comparison as function of time, for the case, load factor $\rho = 1$ and $b_1 = 200$, k_1=5000, $b_2 = 120$, $k_2 = 2000$, $b_3 = 40$, $k_3 = 500$.

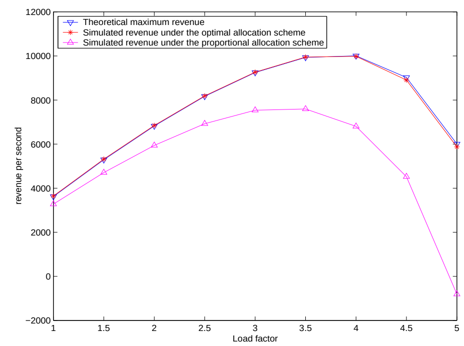

Fig. 7. Revenue comparison as function of load factor ρ, for the case, $b_1 = 200$, k_1=5000, $b_2 = 120$, $k_2 = 2000$, $b_3 = 40$, $k_3 = 500$.

V. CONCLUSIONS

In this paper, we link resource allocation scheme with pricing strategies and explore the problem of maximizing the revenue of network providers by resource allocation among multiple service classes under a certain Service-Level-Agreement and a given amount of network resources. A revenue-aware resource allocation scheme is proposed under linear pricing strategy, which has the closed-form solution to the optimal resource allocation for maximizing the revenue per time unit gained in a network node. The optimal allocation scheme is derived from revenue target function by Lagrangian optimization approach. The simulations demonstrated the revenue-maximizing ability of the optimal resource allocation scheme.

In the future work, the issue of revenue maximization under flat pricing strategy will be investigated. Moreover, revenue criterion as the admission control mechanism will be studied.

REFERENCES

[1] T. Basar, R. Srikant, ”Revenue-maximizing rpicing and capacity expansion in a many-users regime,” Proc. of IEEE INFOCOM2002, Vol.1, 2002, pp. 294-301.

[2] Z. Cao, E. W. Zegura, ”Utility Max-Min: An Application-Oriented Bandwidth Allocation Scheme,” IEEE INFOCOM99, New York, USA, 1999.

[3] C. Courcoubetis, F. P. Kelly, and R. Weber, ”Measurement-based usage charges in communication networks,” Oper. Res., Vol.48, no.4, 2000, pp. 535-548.

[4] T. Hämäläinen and J. Joutsensalo, ”Link Allocation and Revenue Optimization for Future Networks”, Proc. of IEEE Globecom 2002, Nov. 2002, Taipei.

[5] J. Joutsensalo and T. Hämäläinen, ”Optimal Link Allocation and Revenue Maximization,” Journal of Communications and Networks, Vol.4, No.2, June 2002, pp. 136-147.

[6] F. P. Kelly, ”Charging and rate control for elastic traffic,” European Transaction on Telecommunication, vol.8, 1997, pp. 33-37.

[7] L. Massoulie, J. Roberts, ”Bandwidth Sharing: Objectives and Algorithms,” IEEE INFOCOM99, New York, USA.

[8] C. Lee, J. Lehoczky, R. Rajkumar, D. Siewiorek, ”On Quality of Service Optimization with Discrete QoS Options,” IEEE Real-Time Technology and Application Symposium, June 1999.

[9] S. H. Low, ”Equilibrium Allocation of Variable Resources for Elastic Traffics,” INFOCOM98, San Francisco, USA, 1998.

[10] I. Ch. Paschalidis and J. N. Tsitsiklis, ”Congestion-dependent pricing of network services,” IEEE/ACM Transactions on Networking, vol.8, April 2000, pp. 171-184.

[11] I. Ch. Paschalidis and Yong Liu, ”Pricing in multiservice loss networks: static pricing, asymptotic optimality and demand subsitution effects,” IEEE/ACM Transactions on Networking, vol.10, Issue: 3, June 2002, pp. 425-438.

[12] R. Rajkumar, C. Lee, J. Lehoczky, D. Siewiorek, ”A Resource Allocation Model for QoS Management,” IEEE Real-Time Systems Symposium, December 1997.

[13] S. Sarkar, L. Tassiulas, ”Fair Allocation of Utilities in Multirate Multicast Networks,” Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing, Urbana, Illinois, USA, 1999.

[14] S. Shenker, ”Fundamental Design Issues for the Future Internet,” IEEE Journal on Selected Areas in Telecommunications, Vol.13, No.7, September 1995.