# REVENUE-AWARE RESOURCE ALLOCATION IN THE FUTURE MULTI-SERVICE IP NETWORKS

Jian Zhang, Timo Hamalainen, Jyrki Joutsensalo
*Dept. of Mathematical Information Technology, University of Jyvaskyla, FIN-40014 Jyvaskyla, Finland, Email: zhang@cc.jyu.fi, timoh@mit.jyu.fi, jyrkij@mit.jyu.fi*

**Abstract:** In the future IP networks, a wide range of different service classes must be supported in a network node and different classes of customers will pay different prices for their used node resources based on their Service-Level-Agreements. In this paper, we link the resource allocation issue with pricing strategies and explore the problem of maximizing the revenue of service providers in a network node by optimally allocating a given amount of node resources among multiple service classes. Under the linear pricing strategy, the optimal resource allocation scheme is derived for the case that no firm Quality-of-Service (QoS) guarantees are required for all service classes, which can achieve the maximum revenue in a network node; moreover, the suboptimal allocation scheme is proposed for the case that all classes have their firm QoS (mean delay) requirements, which can satisfy those required QoS guarantees while still being able to achieve very high revenue close to the analytic maximum one.

## 1. INTRODUCTION

Resource allocation in the multi-service 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, however, the Internet is changing. New sophisticated real-time applications (video conferencing, video on demand, distance learning, etc) require a better and more reliable network performance. Moreover, these applications require firm performance guarantees from the network where certain resources should be reserved for them. The problem of optimal resource allocation for satisfying an end-to-end Quality-of-Service (QoS) re-

quirement in a network of schedulers is usually addressed by partitioning the end-to-end QoS requirement into local requirements and then solving them in each individual node[5,7]. Hence, the problem of further mapping traffics' local QoS requirements (delay in particular) to their resource allocations in a single network node is of practical importance. On the other hand, in the future multi-class Internet, users will have to pay the network service providers for their used resources based on pricing strategies agreed upon in their Service-Level-Agreements (SLA). From the service providers' point of view, the optimal resource allocation scheme for their revenue maximization is very desirable. In this paper, we addressed the problem of optimizing the resource allocation in a network node for satisfying both the local QoS requirements of multi-class traffics and the revenue maximization of service providers.

Pricing research in the network has been quite intensive during the past few years[2,9,10]. Also a lot of work[12,8,4] has been done concerning the issues of resource allocation and fairness in a single-service environment. The combination of pricing strategies and resource allocation among multiple service classes have not been analyzed widely. A number of works[1,11,6] recently use end-users' utility as the maximizing objective for resource allocation schemes. All of these approaches have a common objective in maximizing the network performance in terms of the users' utility. Our research differs from these studies by linking the resource allocation scheme together with certain pricing strategies to maximize the revenue of a service provider under a given amount of resources.

This paper extends our previous QoS and pricing research[3] and addresses the problem of revenue maximization in a network node by novel revenue-aware resource allocation schemes. Specifically, in a network node supporting multiple service classes, packets are queued in a multi-queue system, where each queue corresponds to one service class. The service provider will receive certain revenues or suffer certain penalties based on given pricing strategies whenever serving a packet. For the case that the service classes supported in a network node are all delay-insensitive, i.e., all service classes have no firm QoS (mean delay, in this paper) requirements, we derive the optimal resource allocation scheme by which the maximum revenue can be obtained. Moreover, when the service classes supported in a network node are all delay-sensitive, i.e., firm QoS (mean delay) guarantees are required for all classes, a suboptimal resource allocation scheme is proposed, which can satisfy those local QoS guarantees while still achieving very high revenue. The simulation results demonstrated the performances of our proposed revenue-aware resource allocation schemes.

The rest of the paper is organized as follows. In Section 2, the linear pricing strategy, which is used in this paper, is generally defined. Revenue-aware resource allocation is investigated in Section 3, where optimal/suboptimal al-

location schemes are derived. Section 4 contains the simulation part evaluating the performances of our proposed resource allocation schemes. Finally, in Section 5, we present our concluding remarks.

## 2. PRICING STRATEGY

As we know, linear, flat and piecewise linear strategies are believed to the most used one in practice. In this paper, our study concentrates on the revenue-maximizing issue under the linear pricing strategy and the analysis under the 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. First some parameters and notions are defined. We consider a network node which supports multiple service classes. Here, 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 amongst 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 QoS considered in this paper focuses on *packet delay*. We use $d_i$ to denote class $i$ packet delay in the network node. For each service class, a pricing function $r_i(d_i)$ is defined to rule the relationship between the QoS (packet delay here) offered to class $i$ customers and the price which class $i$ customers should pay for that QoS. Obviously, it is non-increasing with respect to the delay $d_i$. Specifically, the linear pricing strategy for class $i$ is characterized by the following function.

**Definition 1**: *The function*

$$r_i(d_i) = b_i - k_i d_i, \qquad i = 1, 2, ..., m, b_i > 0, k_i > 0 \tag{1}$$

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 a higher priority than class $j$ (in this paper, we assume that class 1 is the highest priority and class $m$ is the lowest one).

From Eq. (1), it is observed that the constant shift $b_i$ determines the maximum price paid by class $i$ customers and the growing rate of penalty paid to class $i$ with delay depends on the slope $k_i$. Clearly, the set of eligible pricing functions should show that the constant shift $b_1$ and the slope $k_1$ from the highest priority class are both maximal. This is actually what we expect based on the requirements of the Service-Level-Agreement.

# 3. REVENUE-AWARE RESOURCE ALLOCATION

In this section, we consider a network node with capacity $C$ bits/s, supporting $m$ service classes totally. In the node, each class has its own queue. Assume that the queues corresponding to different classes are infinite in length and the packets in the same queue are served in the order they arrive. As most traffic arrival processes have been proven to be Poisson process, the used source traffic model for each class consists of Poisson arrivals and an exponential packet length distribution in this paper. The arrival rate for the $m$ classes is $\lambda_1$,$\lambda_2$,...,$\lambda_m$ (packets/s), respectively. We use $\bar{L}_i$ to denote the mean packet length (in bits) of class $i$. As mentioned above, $d_i$ is used to denote class $i$ packet delay in the node, which consists of the waiting time in queue $i$ and the service time. The share of node capacity allotted to class $i$ is specified by parameter $w_i$, which is called the weight of class $i$. Obviously, the constraints for $w_i$, $1 \leq i \leq m$ are $\sum_{i=1}^{m} w_i = 1$ and $w_i \in (0, 1]$. As a necessary stability condition, $\sum_{i=1}^{m} \lambda_i \bar{L}_i < C$ is required.

As class $i$ packets arrive at queue $i$ with rate $\lambda_i$ and they are guaranteed to receive a portion of node capacity $w_i C$, the analytic mean packet delay $\hat{\bar{d}}_i$ of class $i$ in the node can be estimated as $\hat{\bar{d}}_i = \frac{1}{\frac{w_i C}{\bar{L}_i} - \lambda_i} = \frac{\bar{L}_i}{w_i C - \lambda_i \bar{L}_i}$ based on the queuing theory. Its natural constraint is $w_i C > \lambda_i \bar{L}_i$ due to the fact that delay can not be negative.

Note that a service provider will obtain a revenue or penalty whenever serving one packet. Hence, the metric of revenue used in this paper is the revenue gained per time unit by a service provider. Unless stated otherwise, we shall hereafter refer to the revenue per time unit as revenue. We use the above analytic mean packet delay $\hat{\bar{d}}_i$ to estimate class $i$ packet delay $d_i$. Then the revenue $F$ gained by a service provider in the node may be defined as follows when the linear pricing function in Eq. (1) is deployed:

$$F = \sum_{i=1}^{m} \lambda_i r_i(d_i) = \sum_{i=1}^{m} \lambda_i (b_i - \frac{k_i \bar{L}_i}{w_i C - \lambda_i \bar{L}_i}) \tag{2}$$

## 3.1 Case 1: the resource allocation scheme when no firm QoS guarantees are required

In this subsection, we derive the optimal resource allocation scheme for the case that no firm QoS (delay) guarantees are required for all classes. In other words, packet delay $d_i$ of class $i$ can be any positive value in this case. Then, the issue of revenue maximization in a network node can be formulated as

follows based on the revenue definition in Eq. (2):

$$max \qquad F = \sum_{i=1}^{m} \lambda_i (b_i - \frac{k_i \bar{L_i}}{w_i C - \lambda_i \bar{L_i}}) \tag{3}$$

$$s.t. \qquad \sum_{i=1}^{m} w_i = 1, \;\; 0 < w_i \leq 1 \tag{4}$$

$$w_i C > \lambda_i \bar{L_i} \tag{5}$$

**Theorem 1.** *When no firm delay guarantees are required for all classes, the globally maximum revenue F obtained in a network node is achieved by using the following optimal resource allocation scheme*:

$$w_i = \frac{\sqrt{\lambda_i k_i \bar{L_i}}(C + \frac{\sum_{j=1}^{m} \sqrt{\lambda_j k_j \bar{L_j}}}{\sqrt{\lambda_i k_i \bar{L_i}}} \lambda_i \bar{L_i} - \sum_{j=1}^{m} \lambda_j \bar{L_j})}{C \sum_{j=1}^{m} \sqrt{\lambda_j k_j \bar{L_j}}} \tag{6}$$

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

**Proof:** Based on Eqs. (3) and (4), we can construct the following Lagrangian equation.

$$P = P(w_1, w_2, ..., w_m) = \sum_{i=1}^{m} \lambda_i (b_i - \frac{k_i \bar{L_i} d_0}{w_i - \lambda_i \bar{L_i} d_0}) + \sigma(1 - \sum_{i=1}^{m} w_i) \tag{7}$$

Set partial derivatives of $P$ in Eq. (7) to zero, i.e., $\frac{\partial P}{\partial w_i} = \frac{\lambda_i k_i \bar{L_i} C}{(w_i C - \lambda_i \bar{L_i})^2} - \sigma = 0$. It follows that $\sigma = \frac{\lambda_i k_i \bar{L_i} C}{(w_i C - \lambda_i \bar{L_i})^2}$, leading to the solution:

$$w_i = \sqrt{\frac{\lambda_i k_i \bar{L_i}}{C\sigma}} + \frac{\lambda_i \bar{L_i}}{C}, \;\; i = 1, 2, ..., m. \tag{8}$$

Substituting Eq. (8) to Eq. (4), we get $\sqrt{\sigma} = \frac{\sum_{i=1}^{m} \sqrt{\lambda_i k_i \bar{L_i} C}}{C - \sum_{i=1}^{m} \lambda_i \bar{L_i}}$ and when this $\sqrt{\sigma}$ is substituted back to Eq. (8), the closed-form solution in Eq. (6) is obtained.

Due to the constraint $w_i C > \lambda_i \bar{L_i}$ in (5), obviously, $\sum_{j=1}^{m} w_j C = C > \sum_{j=1}^{m} \lambda_j \bar{L_j}$. Hence, the closed-form solution in Eq. (6) $w_i > 0$. Moreover, this inequality holds: $\lambda_i \bar{L_i} - \frac{\sqrt{\lambda_i k_i \bar{L_i}} \sum_{j \neq i, j=1}^{m} \lambda_j \bar{L_j}}{\sum_{j \neq i, j=1}^{m} \sqrt{\lambda_j k_j \bar{L_j}}} \leq C$, leading to in Eq. (6) the numerator less than the denominator and thus $w_i < 1$. Hence, we can conclude that the closed-form solution in Eq. (6) $w_i \in (0, 1]$ and it is an eligible weight.

To prove that the closed-form solution in Eq. (6) is the only optimal one in the interval (0, 1], we consider the second order derivative of $P$: $\frac{\partial^2 P}{\partial w_i^2} =$

$-\frac{2\lambda_i k_i \bar{L}_i c^2}{(w_i C-\lambda_i \bar{L}_i)^3} < 0$ due to the constraint $w_i C > \lambda_i \bar{L}_i$ in (5). Therefore, the revenue $F$ is strictly convex with the allotted set of weights $\{w_1, ..., w_i, ..., w_m\}$ for the interval $0 < w_i \leq 1$, having one and only one maximum. Hence, the closed-form solution $w_i$ in Eq. (6) is the optimal weight of class *i*, *i*=1,2,...,m. This completes the proof. **Q.E.D.**

Furthermore, the maximum revenue obtained in a network node can be calculated as follows.

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

$$F_{max} = \sum_{i=1}^{m}(\lambda_i b_i) - \frac{(\sum_{i=1}^{m}\sqrt{\lambda_i k_i \bar{L}_i})^2}{C - \sum_{i=1}^{m}\lambda_i \bar{L}_i} \tag{9}$$

**Proof:** Substituting the optimal weight in Eq. (6) to Eq. (2), the maximum revenue $F_{max}$ which can be obtained in a network node is

$$\begin{aligned} F_{max} &= \sum_{i=1}^{m}(\lambda_i b_i - \frac{\lambda_i k_i \bar{L}_i \sum_{i=1}^{m}\sqrt{\lambda_i k_i \bar{L}_i}}{\sqrt{\lambda_i k_i \bar{L}_i}(C - \sum_{i=1}^{m}\lambda_i \bar{L}_i)}) = \\ &= \sum_{i=1}^{m}(\lambda_i b_i) - \frac{(\sum_{i=1}^{m}\sqrt{\lambda_i k_i \bar{L}_i})^2}{C - \sum_{i=1}^{m}\lambda_i \bar{L}_i} \end{aligned} \tag{10}$$

**Q.E.D.**

## 3.2 Case 2: the resource allocation scheme when the firm QoS guarantees are required

In this subsection, we derive the resource allocation scheme in a network node which should satisfy the required QoS guarantees of all classes while still achieving as higher revenue as possible. In this paper, the firm QoS guarantee of class *i* means that the mean packet delay $\bar{d}_i$ of class *i* must be less than the given value $D_i$, i.e., $\bar{d}_i \leq D_i$. We use the analytic mean packet delay $\hat{\bar{d}}_i$ to estimate $\bar{d}_i$, i.e., $\hat{\bar{d}}_i = \frac{\bar{L}_i}{w_i C - \lambda_i \bar{L}_i} \leq D_i$, leading to $w_i \geq \frac{\bar{L}_i}{C}(\lambda_i + \frac{1}{D_i})$. Thus, the required minimum weight of class *i* for satisfying its firm mean delay guarantee is acquired, which we use $w_{i,minimum}$ to denote: $w_{i,minimum} = \frac{\bar{L}_i}{C}(\lambda_i + \frac{1}{D_i})$. Then, the issue of revenue maximization in Case 2 can be formulated as below:

$$max \quad F = \sum_{i=1}^{m}\lambda_i(b_i - \frac{k_i \bar{L}_i}{w_i C - \lambda_i \bar{L}_i}) \tag{11}$$

$$s.t. \quad \sum_{i=1}^{m} w_i = 1, \quad 0 < w_i \leq 1 \tag{12}$$

$$w_i \geq w_{i,minimum} \tag{13}$$

Note that the constraint in (5) has been contained in (12).

To address the above issue, we first calculate the optimal solution by Eq. (6) and it is referred to as $w_{i,optimal}, i = 1, 2, ..., m$ hereafter. Then, if the inequality $w_{i,optimal} > w_{i,minimum}$ holds for all classes, then the optimal solution $w_{i,optimal}, i = 1, 2, ..., m$ by Eq. (6) is the optimal resource allocation scheme in this case because not only can it achieve the maximum revenue but it can also satisfy the firm mean delay guarantees.

If the inequality ($w_{i,optimal} > w_{i,minimum}$) can not hold for all classes, the optimal solution $w_{i,optimal}$ is not eligible to be the resource allocation scheme in this scenario as it cannot satisfy all the firm mean delay guarantees. Additionally, as the constraint in (12) is derived based on the analytic mean delay $\hat{\bar{d}}_i$, the weight allotted to class $i$ in this scenario should actually satisfy the following inequality in (13) to ensure the fulfilment of the firm QoS guarantee of class $i$: $\bar{d}_i \leq D_i$.

$$w_i \geq w_{i,minimum} + \epsilon_i \tag{14}$$

where $\epsilon_i$ is a small positive constant and may be set at different values depending on the accuracy of the implemented firm QoS guarantee. In other words, there is no optimal resource allocation scheme in this scenario. However, as Theorem 1 shows that the revenue $F$ is strictly convex to the allotted set of weights and only has one maximum, the suboptimal resource allocation scheme can be derived by having the suboptimal weight ($w_{i,suboptimal}$) as near $w_{i,optimal}$ as possible and meanwhile satisfying the constraints in Eqs. (11) and (13). We propose a feasible approach to derive the suboptimal allocation scheme for this scenario below.

The value of ($w_{i,optimal} - w_{i,minimum}$) is defined as the weight distance of class $i$. First, all $w_{i,minimum}$ ( $i$=1,2,...,m, $i$ is class index) are sorted by the defined weight distance in descending order so that we acquire a new series of weights denoted by $w_{p,minimum}$ ($p$=1,2,...,m, $p$ is position index). Obviously, there is one mapping between class index $i$ and position index $p$. Next, all $w_{i,minimum}$ are sorted by size also in descending order to achieve another series of weights denoted by $w'_{p,minimum}$ ($p$=1,2,...,m). Then, the suboptimal weight is derived as follows:

$$w_{p,suboptimal} = w_{p,minimum} + \frac{w'_{p,minimum}}{\sum_{p=1}^{m} w_{p,minimum}} (1 - \sum_{p=1}^{m} w_{p,minimum}) \tag{15}$$

for $p = 1, 2, ..., m$. As mentioned above, there is the mapping between class index $i$ and position index $p$, hence, the suboptimal weight $w_{i,suboptimal}$ of class $i$ can be acquired by $w_{p,suboptimal}$.

# 4. SIMULATIONS AND RESULTS

In this section we present some simulation results to illustrate the effectiveness of our revenue-aware resource allocation scheme for maximizing the revenue of service providers while also satisfying the firm mean delay guarantees if needed. A number of simulations have been conducted under different parameter settings. A representative set of these simulations are presented herein. In accord with the presentation format in Section 3, this section also consists of two subsections, one for the case that no mean delay guarantees are required in the simulations and the other one for the case that the supported service classes all have the settings of their firm mean delay guarantees in the simulations. The configuration of all simulations in this section is described below.

Throughout this section, we shall focus on a network node with capacity $C = 10^6$ bits/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 as follows: for the gold class, $\lambda_1$=10 packet/s, for the silver class, $\lambda_2$=15 packet/s, for the bronze class, $\lambda_3$=20 packet/s, and $\bar{L}_i$=3360 bits, $i$=1,2,3. 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 the class-$j$ arrival rate. The set of linear pricing functions deployed is $\{r_1(d_1) = 200 - 10d_1, r_2(d_2) = 150 - 5d_2, r_3(d_3) = 80 - 2d_3\}$ (the time unit of delay is *ms* here).

## 4.1 Case 1 simulations

In this case, the simulations were made for evaluating the performance of the optimal resource allocation scheme derived by Theorem 1. The simulation-generated revenue value by the optimal allocation scheme will be compared with the maximum revenue value calculated by Theorem 2, which is called the analytic maximum revenue value hereafter. In the simulations, the proportional resource allocation scheme, which proportionally allocates the resource amongst all service classes, is also employed for comparison. Specifically, the *proportional* scheme allots the weight of a 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.

Furthermore, we first investigate the evolution of the simulation-generated revenue by the optimal scheme with the time. In this scenario, only the above base arrival rates were used, i.e., load factor $\rho$=1. Additionally, a set of given weights ($w_1$=0.60, $w_2$=0.25, $w_3$=0.15) was also employed for further illustrating the performance of the optimal allocation scheme. Figure 1(a) presents the simulation results when the above set of linear pricing functions is used, where

the x-axis represents the time (the measurement period is 100 seconds here) and the y-axis represents the revenue.

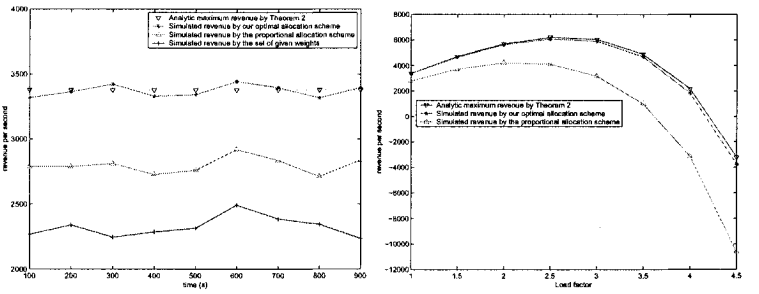
It is observed from Figure 1(a) that the largest simulation-generated revenue value is always achieved by the optimal allocation scheme compared to those by the proportional scheme and the given set of weights. Moreover, the simulation-generated revenue value by the optimal scheme is always quite close to the analytic maximum revenue value, which demonstrates the effectiveness of Theorem 1 and 2. As the parameters in Eq. (9) for calculating the analytic maximum revenue are invariable with time in this scenario, the analytic maximum revenue value does not change with time in Figure 1(a); whereas, as the simulation-generated packet delays are variable, the simulation-generated revenue values vary with time in the figure.

Next we evaluate the performance of the optimal allocation scheme when different traffic intensities are fed into the network node. Figure 1(b) shows the simulation results, where the x-axis represents the load factor and the y-axis represents the revenue. It is seen in Figure 1(b) that the optimal allocation scheme achieves the largest revenue in the simulations under all traffic intensities, which is also very close to the analytic maximum revenue. Moreover, both revenue curve grow almost linearly under light and medium loads. 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. Although the revenue curve of the proportional scheme also grows under light loads, it starts to decrease much earlier as the penalties incurred by the proportional scheme are much larger than the ones by the optimal scheme under the same traffic load.

Based on the above simulation results, we can conclude that the Theorem 1 does derive the optimal resource allocation scheme for the case that no firm QoS guarantees are required, which can achieve the maximum revenue under different traffic intensities.

## 4.2 Case 2 simulations

In this subsection, we evaluate the performances of our derived revenue-aware resource allocation schemes for the case that all classes require their firm QoS (mean delay) guarantees. Throughout this subsection, the firm mean delay guarantees for the gold, silver and bronze classes are set as follows: $\bar{d_1} \leq 10ms, \bar{d_2} \leq 15ms, \bar{d_3} \leq 30ms$. According to the analysis in Section 3, the optimal weight $w_{i,optimal}$ and the required minimum weight $w_{i,minimum}$ should first be calculated, respectively, then the optimal or suboptimal resource allocation scheme in this case can be derived based on the comparison of them. Obviously, if the sum of the calculated $w_{i,minimum}$ exceeds 1 for any load




(a) (b)

*Figure 1.* Case 1 simulations: (a) Revenue comparison as function of time (load factor $\rho = 1$); (b) Revenue comparison as function of load factor $\rho$.

*Table 1.* The calculated $w_{i,optimal}$ and $w_{i,minimal}$, $i$=1, 2, 3, for Case 2 simulations

| | $w_{1,opt}$ | $w_{2,opt}$ | $w_{3,opt}$ | $w_{1,min}$ | $w_{2,min}$ | $w_{3,min}$ |
|---|---|---|---|---|---|---|
| $\rho = 1$ | 0.3733 | 0.3446 | 0.2821 | 0.3696 | 0.2744 | 0.1792 |
| $\rho = 1.5$ | 0.3599 | 0.3436 | 0.2965 | 0.3864 | 0.2996 | 0.2128 |
| $\rho = 2$ | 0.3464 | 0.3426 | 0.3110 | 0.4032 | 0.3248 | 0.2464 |

factor, it means that the node capacity is not enough to satisfy those firm QoS guarantees under such load intensity.

Specifically, in this case, $\sum_{i=1}^{3} w_{i,minimum} > 1$ holds for any load factor $\rho$ which is more than 2.5. Hence, the load factors $\rho$=1, $\rho$=1.5 and $\rho$=2 were used to generate the traffic loads in the following simulations. The above set of linear pricing functions is employed in the simulations, then the calculated $w_{i,optimal}$ and $w_{i,minimum}$ for load factor $\rho$=1, $\rho$=1.5 and $\rho$=2 are summarized in Table 1. Based on the analysis in Section 3, we know that, if $w_{i,optimal} > w_{i,minimum}$ holds for all classes under a load intensity, then the optimal resource allocation scheme ($w_{i,optimal}$, $i$=1,2,3) exists for that load intensity. Specifically, Table 1 shows that the optimal allocation scheme exists for the traffic intensity: $\rho$=1, and thus the optimal weights ($w_{i,optimal}$ shown in Table 1 were also used in this case simulations for this load factor $\rho$=1.

For the scenarios that $w_{i,optimal} > w_{i,minimum}$ does not hold for all classes, the suboptimal resource allocation scheme has to be derived, which should satisfy the required mean delay guarantees while still being able to achieve high revenue. One feasible approach was proposed in Section 3 to derive such suboptimal allocation scheme. Below we illustrate how to calculate the suboptimal weight by that approach.

From Table 1, we observe that the suboptimal weight has to be derived for these two load intensities: $\rho$=1.5, $\rho$=2. Furthermore, under the above load intensities, it is observed from Table 1 that the weight distance of the gold class ($w_{1,optimal} - w_{1,minimum}$) is always the largest and the one of the bronze class ($w_{3,optimal} - w_{3,minimum}$) the smallest among the weight distances of the three classes; moreover, the required minimum weight of the gold class ($w_{1,minimum}$) is also always the largest and the one of the bronze class the smallest among all $w_{i,minimum}$, $i$=1,2,3. Hence, for this special scenario, the formula in Eq. (14) for deriving the suboptimal weight by our proposed approach can be expressed as follows:

$$w_{i,suboptimal} = w_{i,minimum} + \frac{w_{m+1-i,minimum}}{\sum_{i=1}^{m} w_{i,minimum}}(1 - \sum_{i=1}^{m} w_{i,minimum}) \quad (16)$$

for $i$=1, 2, 3, which exactly shows that the left portion of nodal capacity is distributed among all service classes based on their weight distances. Then, the suboptimal weights ($w_{i,suboptimal}$, $i$=1,2,3) for the above load intensities may be calculated by Eq. (15) and summarized as follows: for $\rho$ = 1.5, $w_{1,suboptimal}$ = 0.4104, $w_{2,suboptimal}$ = 0.3333 and $w_{3,suboptimal}$ = 0.2563; for $\rho$ = 2, $w_{1,suboptimal}$ = 0.4097, $w_{2,suboptimal}$ = 0.3333 and $w_{3,suboptimal}$ = 0.2570.

Additionally, another set of weights denoted by $w_{i,comparison}$, $i$=1,2,3 are also employed in the simulations for the comparison and calculated by equation: $w_{i,comparison} = w_{i,minimum} + \frac{w_{i,minimum}}{\sum_{i=1}^{m} w_{i,minimum}}(1 - \sum_{i=1}^{m} w_{i,minimum})$, $i$=1,2,3, which means that the left portion of nodal capacity is allotted to class $i$ only based on the size of its $w_{i,minimum}$. They are summarized below: for $\rho$ = 1, $w_{1,comparison}$ = 0.4490, $w_{2,comparison}$ = 0.3333 and $w_{3,comparison}$ = 0.2177; for $\rho$ = 1.5, $w_{1,comparison}$ = 0.4299, $w_{2,comparison}$ = 0.3333 and $w_{3,comparison}$ = 0.2368; for $\rho$ = 2, $w_{1,comparison}$ = 0.4138, $w_{2,comparison}$ = 0.3333 and $w_{3,comparison}$ = 0.2529.

Then, the above derived optimal/suboptimal weights and the above comparison weights were used in the simulations, respectively. Figure 2 presents the simulation results, which shows that the simulation-generated revenue by the optimal/suboptimal weights is always higher than the one by the comparison weights and it is also pretty close to the analytic maximum revenue.

Moreover, the simulation-generated mean packet delay by the optimal/suboptimal allocation schemes are: 9.9882 *ms*, 11.5756 *ms* and 16.0049 *ms* for $\rho$=1; 9.4695 *ms*, 13.3077 ms and 22.2440 *ms* for $\rho$=1.5; 9.9636 *ms*, 14.8071 *ms* and 28.0530 *ms* for $\rho$=2, which shows that the simulation-generated mean packet delays by the derived optimal/suboptimal scheme satisfy the required firm mean delay guarantees ($\bar{d}_1 \leq 10ms, \bar{d}_2 \leq 15ms, \bar{d}_3 \leq 30ms$). Therefore, it is demonstrated that our derived revenue-aware resource allocation scheme (optimal/suboptimal scheme) in Case 2 is an eligible one, which satisfies all

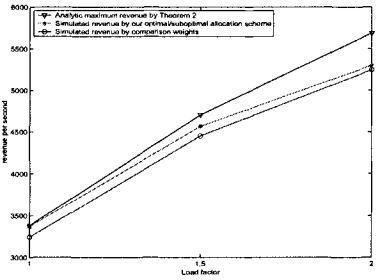



*Figure 2.* Revenue comparison as function of load factor $\rho$ in Case 2 simulations

required firm QoS (mean delay) guarantees while still achieves very high revenue (pretty close to the analytic maximum one).

## 5. CONCLUSIONS

In this paper, we link the resource allocation issue with the pricing strategies and explore the problem of maximizing the revenue of service providers by optimally allocating a given amount of resource among multiple service classes. Under the linear pricing strategy, the optimal allocation scheme is derived for the case that no firm QoS guarantees are required for all classes, which can achieve the maximum revenue in a network node; moreover, the suboptimal allocation scheme is proposed for the case that all classes have their firm QoS (mean delay) requirements, which can satisfy those required QoS guarantees while still being able to achieve very high revenue close to the analytic maximum one. The simulation results demonstrated the effectiveness of our proposed optimal/suboptimal resource allocation schemes.

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

## REFERENCES

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

2. 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.

3. J. Joutsensalo, T. Hamalainen, M. Paakkonen, and A. Sayenko, "Revenue Aware Scheduling Algorithm in The Single Node Case," Journal of Communications and Networks, Vol.6, No.1 March 2004.
4. L. Massoulie, J. Roberts, "Bandwidth Sharing: Objectives and Algorithms," IEEE INFOCOM99, New York, USA.
5. R. Nagarajan, J. F. Kurose and D. Towsley, "Allocation of Local Quality of Service Constraints to Meet End-to-End Requirements," IFIP Workshop on the Performance Analysis of ATM Systems, Martinique, Jan. 1993.
6. 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.
7. D. H. Lorenz and A. Orda, "Optimal partition of QoS requirements on unicast paths and multicast trees," Proceedings of IEEE INFOCOM'99, pp. 246-253, March 1999.
8. S. H. Low, "Equilibrium Allocation of Variable Resources for Elastic Traffics," INFOCOM98, San Francisco, USA, 1998.
9. 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.
10. 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.
11. 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.
12. S. Shenker, "Fundamental Design Issues for the Future Internet," IEEE Journal on Selected Areas in Telecommunications, Vol.13, No.7, September 1995.