QoS- and Revenue Aware Adaptive Scheduling Algorithm

Jyrki Joutsensalo, Timo Hämäläinen, Alexander Sayenko, and Mikko Pääkkönen

Abstract: **In the near future packet networks should support applications which can not predict their traffic requirements in advance, but still have tight quality of service requirements, e.g., guaranteed bandwidth, jitter, and packet loss. These dynamic characteristics mean that the sources can be made to modify their data transfer rates according to network conditions. Depending on the customer's needs, network operator can differentiate incoming connections and handle those in the buffers and the interfaces in different ways. In this paper, dynamic QoS-aware scheduling algorithm is presented and investigated in the single node case. The purpose of the algorithm is — in addition to fair resource sharing to different types of traffic classes with different priorities — to maximize revenue of the service provider. It is derived from the linear type of revenue target function, and closed form globally optimal formula is presented. The method is computationally inexpensive, while still producing maximal revenue. Due to the simplicity of the algorithm, it can operate in the highly nonstationary environments. In addition, it is nonparametric and deterministic in the sense that it uses only the information about the number of users and their traffic classes, not about call density functions or duration distributions. Also, Call Admission Control (CAC) mechanism is used by hypothesis testing.**

Index Terms: **Packet scheduling pricing, revenue maximization, QoS.**

I. INTRODUCTION

Packet scheduling discipline is an important factor of a network node. The choice of the discipline impacts the allocation of restricted network resources among contending flows of the communication network. Network operators can handle resource reservations by using traffic differentiation and design different kind of pricing strategies. The open question still arises: How to put these two issues together. Pricing research in the networks has been quite intensive during the last years and also novel queuing algorithms have been proposed, but combination of them have not been analyzed widely. Next, we will present summary of the recently made pricing work and after that we will highlight the mostly used queuing disciplines.

A smart market charging method for network usage is presented in [1]. This paper studies individual packets' bid for transport while the network only serves packets with bids above a certain (congestion-dependent) cutoff amount. Charges that increase with either realized flow rate or with the share of the network consumed by a traffic flow is studied in [2] and [3]. Packet-based pricing schemes (e.g., [4], [5]) have also been proposed as an incentive for more efficient flow control. The fundamental problem of achieving the system optimum that maximizes the aggregate utility of the users, using only the information available at the end hosts, is studied in [6]. They assume that the users are of elastic traffic and can adjust their rates based on their estimates of network congestion level. Equilibrium properties of bandwidth and buffer allocation schemes are analyzed in [7]. Pricing and link allocation for real-time traffic that requires strict QoS guarantees is studied, e.g., in [8] and [9]. Such QoS guarantees can often be translated into a preset resource amount that has to be allocated to a call at all links in its route through the network. If the resource is bandwidth, this resource amount can be some sort of an effective bandwidth (see, e.g., [10] for a survey of effective bandwidth characterizations and [11] for similar notions in the multiclass case). In this setting, [12], [13] propose the pricing of real-time traffic with QoS requirements, in terms of its effective bandwidth. Their pricing scheme can also be called as static one and it has clear implementation advantages: charges are predictable by end users, evolve in a slower time-scale than congestion phenomena, and no real-time mechanism is needed to communicate tariffs to the users.

There is also several research work done with the game-theoretic models of routing and flow control in communication networks (e.g., [14]–[19]). These papers show conditions for the existence and uniqueness of an equilibrium. This has allowed, in particular, the design of network management policies that induce efficient equilibria [15]. This framework has also been extended to the context of repeated games in which cooperation can be enforced by using policies that penalize users who deviate from the equilibrium [17]. A revenue-maximizing pricing scheme for the service provider is presented in [20]. Thus, a noncooperative (Nash) flow control game is played by the users (followers) in a Stackelberg game where the goal of the leader is to set a price to maximize revenue.

Two well-known scheduling algorithms are the packet-by-packet generalized processor sharing (PGPS) ([21]) and the worstcase fair weighted fair queueing (WF^2Q) ([22]). The WF^2Q has been proposed to eliminate PGPS burstiness problem exhibited in a flow packet departure process. Based on the fluid traffic model, the generalized processor sharing discipline provides the delay and buffer occupancy bounds for guaranteeing the QoS. The delay bound for the PGPS is provided, e.g., in [21], which is equivalent to the weighted fair queuing (WFQ) [23]. As outlined in [22], the departure process resulting from packet assignment by a PGPS server could be bursty. To avoid this problem, a new packet approximation algorithm of the GPS (i.e., WF^2Q) was proposed in [22]. The queueing disciplines such as PGPS and WF^2Q are based on a timestamp mechanism to determine the packet service sequence. The timestamp mechanism for all packets, however, entails implementation complexity. If a fixed length packet is used, the implementation complexity due to the timestamp mechanism can be reduced, in which a round robin discipline such as the weighted round

Manuscript received March 18, 2003; approved for publication by Byung G. Kim, Division III Editor, August 28, 2003.

The authors are with Department of Mathematical Information Technology University of Jyväskylä, Finalnd, email: {jyrkij, timoh}@mit.jyu.fi, {mikpaak, sayenko}@cc.jyu.fi.

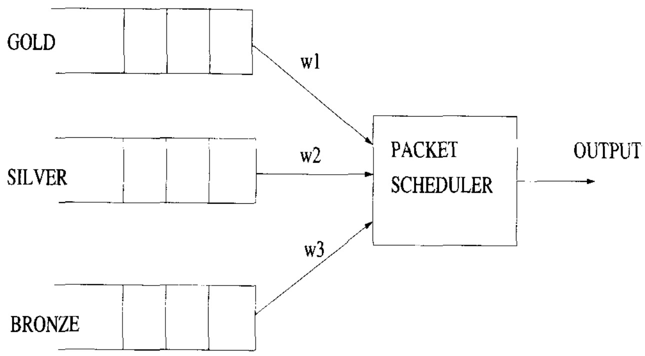


Fig. 1. Traffic classification at the output buffers.

robin (WRR) could be used. Although simple to implement by avoiding the use of timestamp mechanism, the WRR has a larger delay bound. To solve this problem, several modification approaches of the WRR have been proposed. As seen in [24] and [25], the uniform round robin (URR) discipline and the WF^2Q interleaved WRR discipline emulate the WF^2Q to determine the packet service sequence. These scheduling disciplines result in a more uniform packet departure and a smaller delay bound than those provided by conventional round robin. Extension to WRR algorithm for fixed length packets is studied in [26]. They present a scheduling algorithm for fixed length packets that do not emulate the WF^2Q. As the timestamp mechanism is not necessary, the proposed algorithm can be implemented with a low complexity and low processing delay for high speed networks.

Our research differs from the above studies by linking pricing and queuing issues together; in addition our model does not need any additional information about user behavior, utility functions, etc. (like most pricing and game-theoretic ones need). This paper extends our previous pricing and QoS research, [27]–[30], to take into account queuing scheduling issues by introducing dynamic weight tracking algorithm in the scheduler. QoS and revenue aware scheduling algorithm is investigated in the single node case. It is derived from Lagrangian optimization problem, and globally optimal closed form solution is presented.

The rest of the paper is organized as follows. In Section II, used pricing scenario is presented and generally defined. Closed form scheduling algorithm is derived in Section III; in addition, Call Admission Control (CAC) mechanism as well as some upper bounds are presented in this section. Section IV contains experimental part justifying theorems. Discussions are made in Section V, and final section contains conclusions of the work.

II. PRICING SCENARIO

Here the pricing scenario is presented in the simplified form. First, some parameters and notations are defined and commented. Let d_0 be the minimum processing time of the classifier for transmitting data from one queue to the output in Fig. 1. For simplicity it is assumed that the data packets have the same size b. Therefore their size can be scaled to $b = 1$. Extensions to the variable packet sizes do not need essential modifications to the main theory. The number of service classes is denoted by m. Literature usually refers to the gold, silver, and bronze classes;

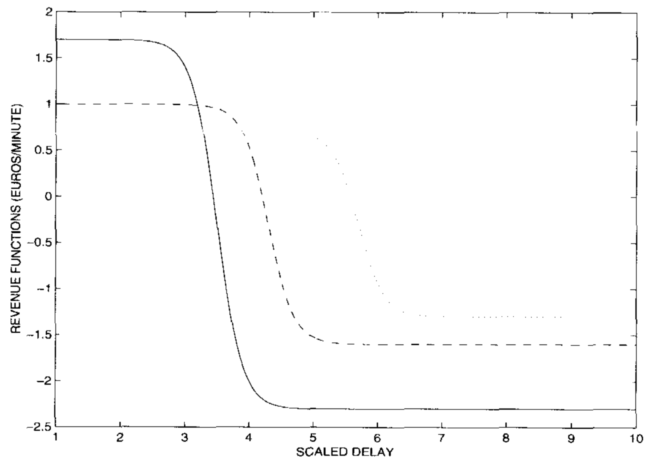


Fig. 2. Three hyperbolic tangent type pricing functions. Horizontal axis: delay. Vertical axis: price.

in this case, $m = 3$. In each queue, sub-queues can be defined due to the different insertion delays, transmission delays, etc. of the different packets in the same queue. However, this is also straightforward extension to our scenario, and therefore it is beyond the scope of this study. It has only the effect on the computational complexity, and is shortly discussed in the Section V-B. Real processing time (delay) for class i in the packet scheduler is

$$d_i = N_i d_0 / w_i, \tag{1}$$

where $w_i(t) = w_i, i = 1, \cdots, m$ are weights allotted for each class, and $N_i(t) = N_i$ is the number of customers in the ith queue. Here time index t has been dropped for convenience. The natural constraint for the weights are

$$w_i > 0, \tag{2}$$

and

$$\sum_{i=1}^{m} w_i = 1. \tag{3}$$

Without loss of generality, only non-empty queues are considered, and therefore $w_i \neq 0, i = 1, \cdots, m$. If some weight is $w_i = 1$, then $m = 1$, and the only class to be served has the minimum processing time d_0, if $N_i = 1$. For each service class, a revenue or **pricing function**

$$r_i(d_i) = r_i(N_i d_0 / w_i + c_i), \tag{4}$$

(euros/minute) is non-increasing with respect to the delay d_i. Here $c_i(t) = c_i$ includes insertion delay, transmission delay, etc., and here it is assumed to be constant (therefore above-mentioned sub-queue systems are not considered here). Examples of pricing functions are given in Figs. 2, 3, and 4. Fig. 2 presents three ($m = 3$) hyberbolic tangent type functions, while Fig. 3 presents piecewise linear pricing functions. A benefit of using hyperbolic tangents is that they are continuously differentiable, and hence adaptive gradient algorithms are quite easy to develop. However, there are some disadvantages: For analytical reasons it is difficult to develop optimal revenue maximization and CAC algorithms. In addition, they are perhaps not tempting from the point view of customers. Piecewise linear pricing

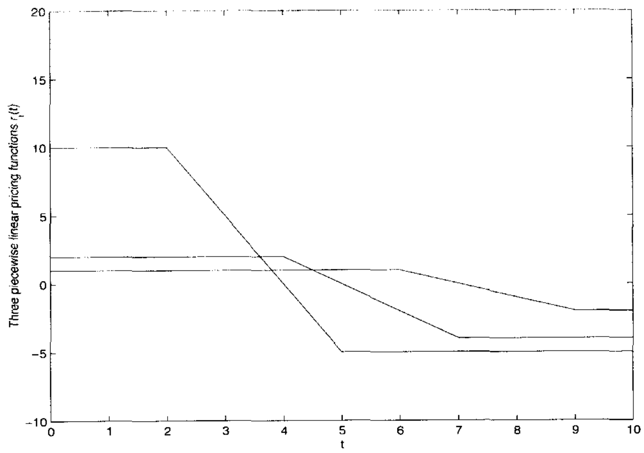


Fig. 3. Three piecewise linear pricing functions. Horizontal axis: delay. Vertical axis: price.

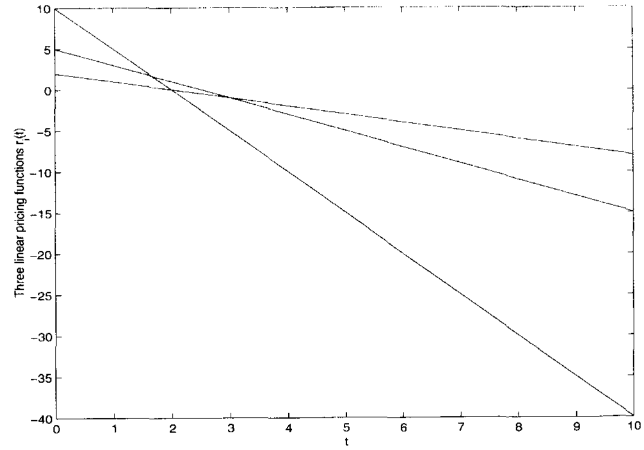


Fig. 4. Three linear pricing functions. Horizontal axis: delay; vertical axis: price.

functions, as shown in Fig. 3, seem to be tempting, and study of those functions are topics of the future research. Our study concentrates on the case of the simplest functions, namely linear pricing functions, as shown in Fig. 4. Linear pricing algorithms may perhaps also be used as building blocks for developing piecewise linear pricing models. Due to the topic, Fig. 4 is commented in more detail. For gold class, the pricing model $r_1(d) = -5d + 10$ in Fig. 4 means that when the delay d is small, the price paid by gold class customer is high - maximally 10 units of money. It is natural that for the highest priority class, constant shift (e.g., ten money units in this case) is selected to be highest. On the other hand, penalty paid to the highest priority class customers is also highest; in this case it depends linearly on the delay, being $-5d$. For example, if $d = 3$, then $r_1(d) = r_1(3) = -5 \times 3 + 10 = -5$ units of money. Same observations hold for silver and bronze classes. For bronze class, $r_3(d) = -d + 2$ means that the price paid by that class customer is maximally 2 units of money. In this case, constant shift was selected to be lowest. On the other hand, penalty for bronze class is also lowest, being $-d$. However, our purpose is not to make accurate study of the practical realizations of the parameters of the curves, only general parametrical forms of the pricing functions.

III. REVENUE MAXIMIZATION ALGORITHM

One user in class i pays $r_i(d_i)$ money units to the service provider according to the pricing function (4). Because there are N_i customers in the queue i, the total price paid by the ith class customers in unit time step (euros/minute) is $N_i r_i(N_i d_0/w_i + c_i)$. Because there are m classes, the revenue criterion to be maximized has the form

$$F(w_1, \cdots, w_m) = \sum_{i=1}^{m} N_i r_i(N_i d_0/w_i + c_i), \tag{5}$$

under weight constraint (2) and (3). Without loss of generality, set $d_0 = 1$.

As a special case, consider *linear* revenue model.

Definition 1: The function

$$r_i(t) = -r_i t + k_i, \quad i = 1, \cdots, m, \tag{6}$$

$$r_i > 0, \tag{7}$$

$$k_i > 0, \tag{8}$$

is called linear pricing function.

Using (4), (5), and (6), we define the revenue F for linear pricing functions by Lagrangian as follows:

$$\begin{aligned} F &= F(w_1, \cdots, w_m) \\ &= \sum_{i=1}^{m} N_i(-r_i \frac{N_i}{w_i} + k_i) + \lambda(1 - \sum_{i=1}^{m} w_i) \\ &= -\sum_{i=1}^{m} \frac{r_i N_i^2}{w_i} + \sum_{i=1}^{m} N_i k_i + \lambda(1 - \sum_{i=1}^{m} w_i), \\ & 0 < w_i \leq 1. \end{aligned} \tag{9}$$

Here the constants c_i have been dropped out for convenience. Theorem of closed form solution for optimal weights is as follows:

Theorem 1: For linear pricing functions, the maximum revenue F is achieved by using the weights

$$w_i = \frac{\sqrt{r_i} N_i}{\sum_{l=1}^{m} \sqrt{r_l} N_l}, \tag{10}$$

and it is unique in $w_i \in (0, 1]$.

Proof: Set partial derivatives of F in (9) to the zero:

$$\frac{\partial F}{\partial w_l} = \frac{r_l N_l^2}{w_l^2} - \lambda = 0. \tag{11}$$

It follows that

$$\lambda = \frac{r_l N_l^2}{w_l^2}, \quad l = 1, \cdots, m, \tag{12}$$

leading to the solution

$$w_i = \frac{\sqrt{r_i} N_i}{\sqrt{\lambda}}, \quad i = 1, \cdots, m. \tag{13}$$

Substituting the scaling factor (3)

$$\sum_{l=1}^{m} w_l = \sum_{l=1}^{m} \frac{\sqrt{r_l} N_l}{\sqrt{\lambda}} = 1, \tag{14}$$

to the denominator of (13), the closed form solution is obtained as follows:

$$w_i = \frac{\frac{\sqrt{r_i} N_i}{\sqrt{\lambda}}}{\sum_{l=1}^{m} \frac{\sqrt{r_l} N_l}{\sqrt{\lambda}}} = \frac{\sqrt{r_i} N_i}{\sum_{l=1}^{m} \sqrt{r_l} N_l}. \tag{15}$$

Because $r_1 > 0$ and $N_i > 0$, it is seen from (15) that

$$\frac{1}{w_i} = 1 + \frac{\sum_{l \neq i}^{m} \sqrt{r_l} N_l}{\sqrt{r_i} N_i} \geq 1. \tag{16}$$

So

$$0 < w_i \leq 1. \tag{17}$$

To prove that this is the only solution in the interval $0 < w_i \leq 1$, we consider second order derivative of F. Due to the constraint $\sum_{l=1}^{m} w_l = 1$, one obtains from (12)

$$\begin{aligned} \lambda &= \sum_{l=1}^{m} \lambda w_l = \sum_{l=1}^{m} \frac{r_l N_l^2}{w_l^2} w_l \\ &= \sum_{l=1}^{m} \frac{r_l N_l^2}{w_l} = -F + \sum_{i=1}^{m} N_i k_i. \end{aligned} \tag{18}$$

The right hand side term follows from the fact that the penalty $\lambda(1 - \sum_{i=1}^{m} w_i)$ vanishes in (9), when the constraint $\sum_{l=1}^{m} w_l = 1$ is satisfied. Solving λ out from (11) using (18), one obtains an expression

$$\frac{\partial F}{\partial w_l} = \frac{r_l N_l^2}{w_l^2} - \lambda = \frac{r_l N_l^2}{w_l^2} - \sum_{l=1}^{m} \frac{r_i N_i^2}{w_i}, \tag{19}$$

to the first order derivative. Then second order derivative is

$$\frac{\partial^2 F}{\partial w_l^2} = r_l N_l^2 \left(-\frac{2}{w_l^3} + \frac{1}{w_l^2} \right) = r_l N_l^2 \frac{-2 + w_l}{w_l^3} < 0, \tag{20}$$

because $r_l > 0$, $N_l > 0$, and $w_l \in (0, 1]$. Therefore, F is strictly concave in the interval $0 < w_i \leq 1$, having one and only one maximum. This completes the proof. □

An upper bound for revenue F is obtained:

Theorem 2:

$$F < \sum_{i=1}^{m} N_i k_i. \tag{21}$$

Proof: Due to the constraint $\sum_{l=1}^{m} w_l = 1$ in (9),

$$F = \sum_{i=1}^{m} N_i k_i - \sum_{i=1}^{m} \frac{r_i N_i^2}{w_i} < \sum_{i=1}^{m} N_i k_i, \tag{22}$$

is obtained, because $r_i > 0$, $N_i > 0$, and $w_i > 0$. □

Analytical form to the revenue can be expressed solving weights w_i out:

Theorem 3: When optimal weights w_i are used according to Theorem 1, revenue is

$$F = -\left(\sum_{i=1}^{m} \sqrt{r_i} N_i \right)^2 + \sum_{i=1}^{m} N_i k_i. \tag{23}$$

Proof: When penalty $\lambda(1 - \sum_i w_i)$ in (9) vanishes, F can be represented in the form

$$F = -\sum_{i=1}^{m} \frac{r_i N_i^2}{w_i} + \sum_{i=1}^{m} N_i k_i. \tag{24}$$

Substitute optimal weights (10) to (24). Then

$$\begin{aligned} F &= -\sum_{i=1}^{m} r_i N_i^2 \frac{\sum_{l=1}^{m} \sqrt{r_l} N_l}{\sqrt{r_i} N_i} + \sum_{i=1}^{m} N_i k_i \\ &= -\sum_{i=1}^{m} \sqrt{r_i} N_i \sum_{l=1}^{m} \sqrt{r_l} N_l + \sum_{i=1}^{m} N_i k_i \\ &= -(\sum_{i=1}^{m} \sqrt{r_i} N_i)^2 + \sum_{i=1}^{m} N_i k_i. \end{aligned} \tag{25}$$

□

From (25), one possible constraint in the CAC mechanism is obtained, namely

$$-(\sum_{i=1}^{m} \sqrt{r_i} N_i)^2 < \sum_{i=1}^{m} N_i k_i, \tag{26}$$

that guarantees $F > 0$.

Next theorem states optimal number of users, as well as upper bounds for buffer sizes:

Theorem 4: Upper bounds for buffer sizes are

$$q_i = \lfloor \frac{1}{2} \frac{k_i}{r_i} \rfloor, \quad i = 1, \cdots, m, \tag{27}$$

where $y = \lfloor x \rfloor$ denotes maximum integer y satisfying $y \leq x$.

Proof: The optimal number of users for fixed weights is obtained as follows:

$$\frac{\partial F}{\partial N_l} = -2 \frac{r_l}{w_l} N_l + k_l = 0. \tag{28}$$

Therefore

$$N_l = \frac{1}{2} \frac{w_l k_l}{r_l}, \quad l = 1, \cdots, m. \tag{29}$$

The second derivative is

$$\frac{\partial^2 F}{\partial N_l^2} = -2 \frac{r_l}{w_l} < 0, \tag{30}$$

because $r_l > 0$ and $w_l \geq 0$. Therefore F is strictly concave with respect to N_i, $i = 1, \cdots, m$ having one and only one global maximum, which is satisfied by (29). Because $w_i \leq 1$, $i = 1, \cdots, m$, then

$$N_l \leq \frac{1}{2} \frac{k_l}{r_l}, \tag{31}$$

for which (27) follows. This completes proof. □

Next another upper bound for revenue is presented:

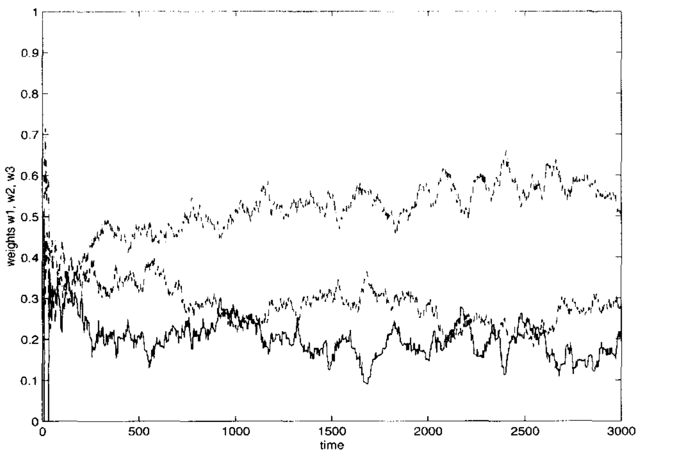


Fig. 5. First experiment with no call admission control. Three weights as a function of time. Horizontal axis: time. Vertical axis: weight value.

Theorem 5: In the case of linear pricing model (4), upper bound for revenue is

$$F \leq \frac{1}{4}\sum_{i=1}^{m}\frac{k_i^2}{r_i}. \tag{32}$$

Proof: Select optimal value for N_i in (29), and substitute it in (9) by using constraint (3). Then

$$F = \sum_{i=1}^{m}\frac{1}{2}\frac{w_i k_i}{r_i}\left(-r_i\frac{1}{2}\frac{w_i k_i}{r_i w_i}+k_i\right) = \frac{1}{4}\sum_{i=1}^{m}\frac{w_i k_i^2}{r_i}. \tag{33}$$

Due to the condition $w_i \leq 1$, (32) follows. □

Interpretation of (32) is quite obvious: k_i increases upper limit, while r_i decreases it.

CAC mechanism can be made by simple hypothesis testing without assumptions about call or dropping rates. Let the state (number of packets) at the moment t be $N_i(t)$, $t = 1, \cdots, m$. Let the new hypothetical state at the moment $t+1$ be $\tilde{N}_i(t+1)$, $i = 1, \cdots, m$, when one or several calls appear. In hypothesis testing, Theorem 3 is applied as follows:

$$F(t) = -\left(\sum_{i=1}^{m}\sqrt{r_i}N_i(t)\right)^2 + \sum_{i=1}^{m}N_i(t)k_i. \tag{34}$$

$$\tilde{F}(t+1) = -\left(\sum_{i=1}^{m}\sqrt{r_i}\tilde{N}_i(t+1)\right)^2 + \sum_{i=1}^{m}\tilde{N}_i(t+1)k_i. \tag{35}$$

If $F(t) > \tilde{F}(t)$, then call is rejected, otherwise it is accepted.

Computational complexity of the algorithm can also be derived by exploiting Theorem 3. When no calls or droppings happen, weights are not adjusted. When call appears, $O(m)$ multiplications and additions are performed, as seen from (23).

IV. EXPERIMENTS

In all the experiments, calls and durations are Poisson and exponentially distributed, respectively. In addtition, number of classes is $m = 3$. Call rates per unit time for gold, silver, and bronze classes are $\alpha_1 = 0.1$, $\alpha_2 = 0.2$, and $\alpha_3 = 0.3$, respectively. Duration parameters (decay rates) are $\beta_1 = 0.010$,

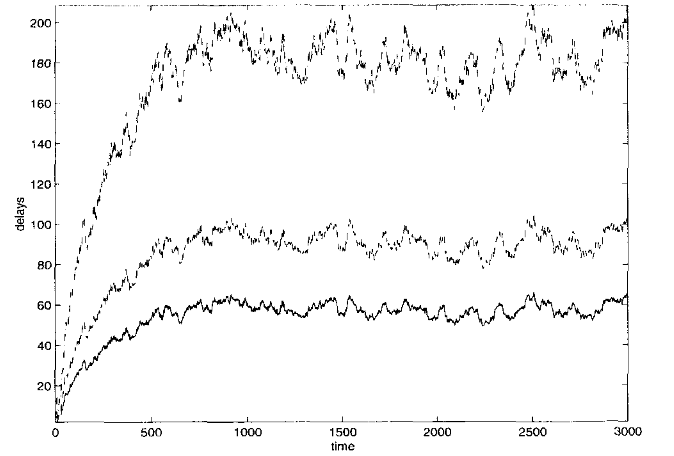


Fig. 6. First experiment with no call admission control. Delays as a function of time. Horizontal axis: time. Vertical axis: delay. Solid, dashed, and dash-dotted curves correspond to gold, silver, and bronze class, respectively.

$\beta_2 = 0.007$, and $\beta_3 = 0.003$, where probability density functions for durations are

$$f_i(t) = \beta_i e^{-\beta_i t}, \quad i = 1, 2, 3, \quad t \geq 0. \tag{36}$$

The number of unit times in the experiments was $T = 3000$.

Experiment 1. In the first experiment, three service classes have the pricing functions

$$r_1(t) = -5t + 200, \tag{37}$$

for gold class,

$$r_2(t) = -2t + 100, \tag{38}$$

for for silver class, and

$$r_3(t) = -0.5t + 50, \tag{39}$$

for bronze class. Fig. 5 shows the evolutions of three weights $w_1(t)$, $w_2(t)$, and $w_3(t)$ as a function of time, when no CAC mechanism is used. Fig. 6 shows the corresponding delays. Solid, dashed, and dash-dotted curves correspond to the gold, silver, and bronze class, respectively. It is not surprising that the delays of gold class customers are lowest, while delays of bronze class customers are largest. Number of users $N_i(t)$ are shown in Fig. 7. Due to the arrival and duration rates, number of users is lowest in gold class, while number of users is largest in bronze class. Solid, dashed, and dash-dotted lines show upper bounds of the different buffers according to the Theorem 4. However, because CAC mechanism is not used, $N_i(t)$ may be larger than the upper bounds. It is seen that $N_i(t)$ achieves the theoretical value

$$E[N_i(t)] = \frac{\alpha_i}{\beta_i}, \tag{40}$$

stated in Little's Theorem, i.e., $\alpha_1/\beta_1 = 10$, $\alpha_2/\beta_2 \approx 29$, and $\alpha_3/\beta_3 = 100$. In Fig. 8, revenue as well as two upper bounds are shown. Lowest curve (solid) is the revenue achieved by the closed form method with no CAC. It becames negative. Dashed curve shows the upper limit $\sum_i N_i k_i$ as stated in Theorem 2, and the solid line illustrates the upper bound of Theorem 5, and

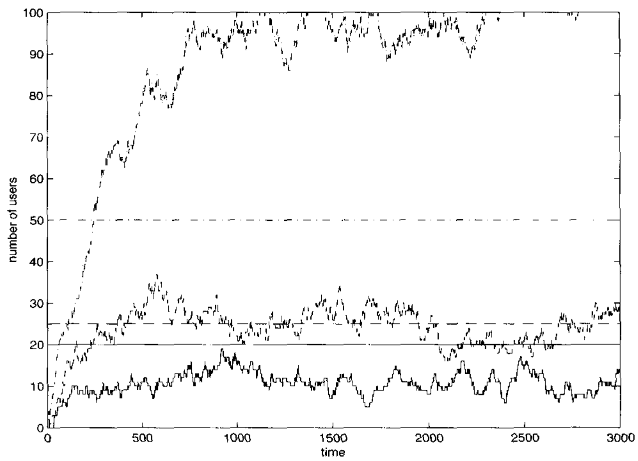


Fig. 7. First experiment with no call admission control. Number of users as a function of time. Horizontal axis: time. Vertical axis: number of users. Solid, dashed, and dash-dotted curves correspond to gold, silver, and bronze class, respectively. Lines are upper bounds for buffer sizes.

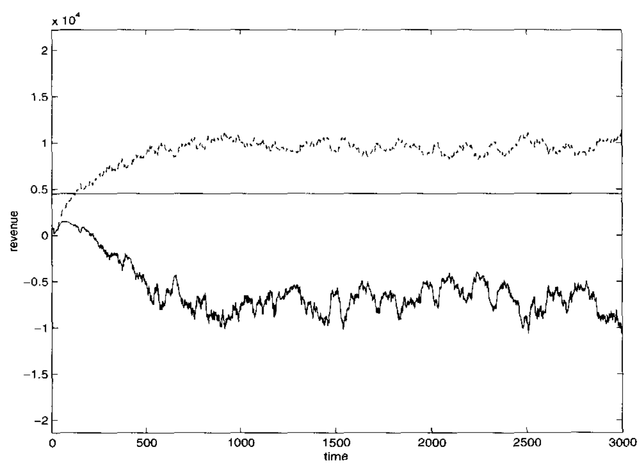


Fig. 8. First experiment with no call admission control. Revenue as a function of time. Horizontal axis: time. Vertical axis: revenue. Lowest curve: realized revenue; middle curve (dashed): upper bound $\sum_i N_i k_i$; solid line: $0.25 \sum_i k_i^2 / r_i$.

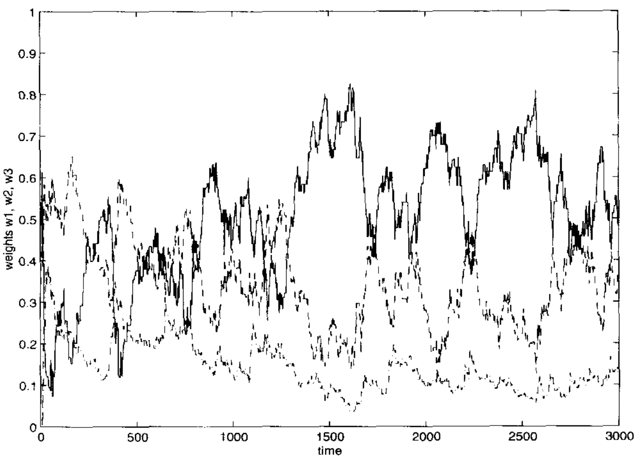


Fig. 9. First experiment with call admission control. Three weights as a function of time. Horizontal axis: time. Vertical axis: weight value.

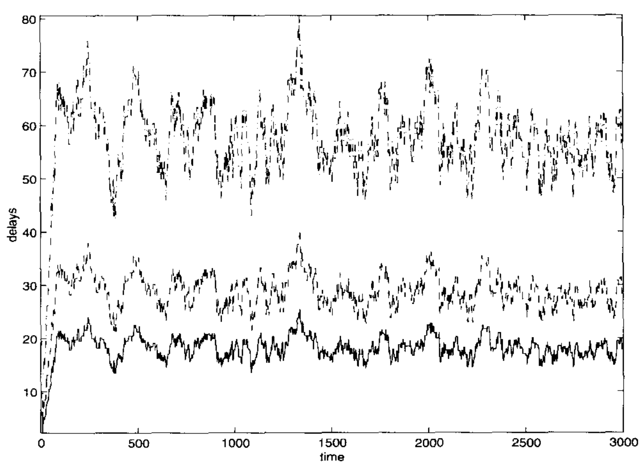


Fig. 10. First experiment with call admission control. Delays as a function of time. Horizontal axis: time. Vertical axis: delay. Solid, dashed, and dash-dotted curves correspond to gold, silver, and bronze class, respectively.

it is constant due to the invariance of r_i and k_i in that experiment. The experiment is continued by using CAC mechanism. The parameters remain the same. Figs. 9–12 show weights, delays, number of users, and revenues, respectively. It is noticeable that the upper bounds are larger than the realized number of users, as well as revenue, and what is important is that the revenue is now positive, as seen from Fig. 12.

Experiment 2. In the second experiment, the pricing functions are

$$r_1(t) = -10t + 200, \tag{41}$$

for gold class,

$$r_2(t) = -4t + 100, \tag{42}$$

for for silver class, and

$$r_3(t) = -t + 50, \tag{43}$$

for bronze class, i.e., penalty coefficients r_i are larger compared to those in the first experiment. Figs. 13 and 14 show number of users and revenue, respectively, for the algorithm that does not use CAC mechanism. The traffic profiles remain the same, i.e., the number of users $N_i(t)$ are the same due to the lack of the CAC mechanism, as seen by comparing Figs. 7 and 13. Revenue remains negative, and it is smaller than in the experiment 1, as seen by comparing Figs. 8 and 14. The reason that the penalty factors r_i are now larger. Figs. 15 and 16, show number of users, and revenue, respectively, for the algorithm that does use CAC mechanism. Number of users as well as revenue are below upper limits, and revenue becomes smaller, as seen by comparing solid curves of Figs. 12 and 16.

Experiment 3. In this experiment, the pricing functions are

$$r_1(t) = -5t + 2000, \tag{44}$$

for gold class,

$$r_2(t) = -2t + 1000, \tag{45}$$

for for silver class, and

$$r_3(t) = -0.5t + 400, \tag{46}$$

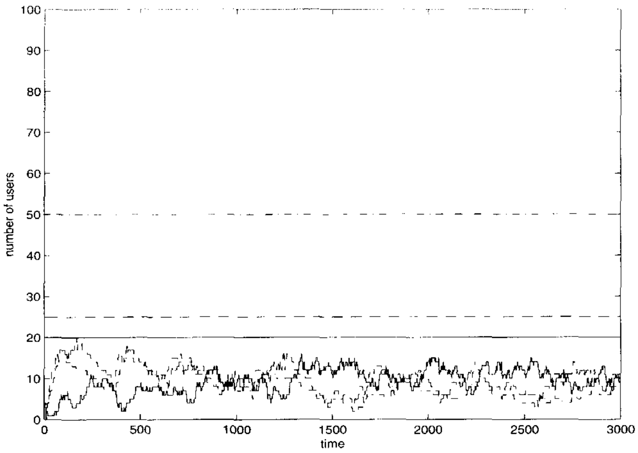


Fig. 11. First experiment with call admission control. Number of users as a function of time. Horizontal axis: time. Vertical axis: number of users. Solid, dashed, and dash-dotted curves correspond to gold, silver, and bronze class, respectively. Lines are upper bounds for buffer sizes.

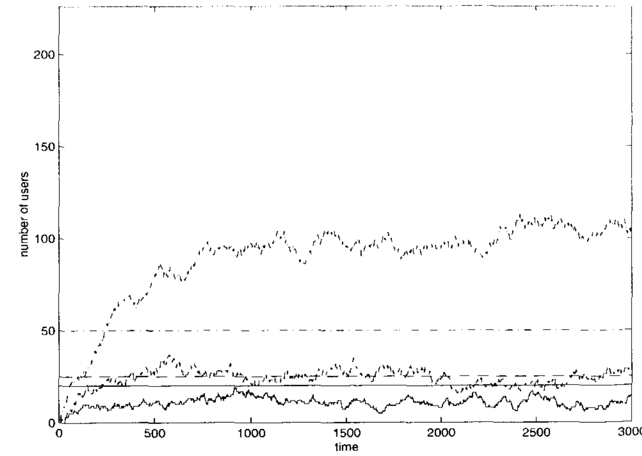


Fig. 13. Second experiment with no call admission control. Number of users as a function of time. Horizontal axis: time. Vertical axis: number of users. Solid, dashed, and dash-dotted curves correspond to gold, silver, and bronze class, respectively. Lines are upper bounds for buffer sizes.

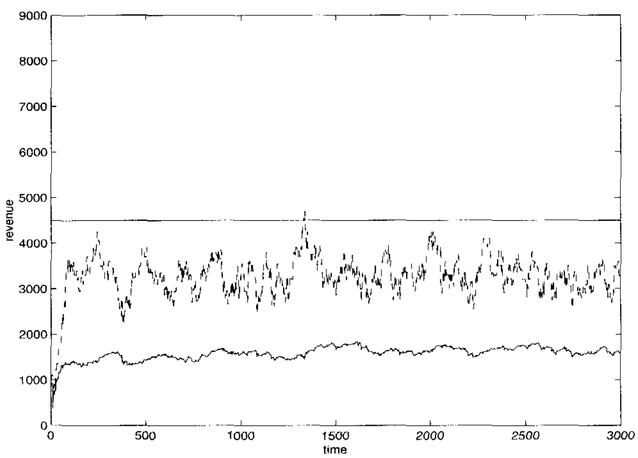


Fig. 12. First experiment with call admission control. Revenue as a function of time. Horizontal axis: time. Vertical axis: revenue. Lowest curve: realized revenue; middle curve (dashed): upper bound $\sum_i N_i k_i$; solid line: $0.25 \sum_i k_i^2/r_i$.

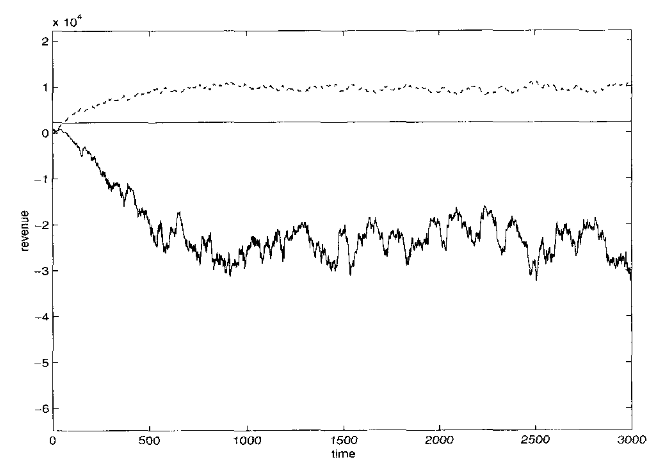


Fig. 14. Second experiment with no call admission control. Revenue as a function of time. Horizontal axis: time. Vertical axis: revenue. Lowest curve: realized revenue; middle curve (dashed): upper bound $\sum_i N_i k_i$; solid line: $0.25 \sum_i k_i^2/r_i$.

for bronze class, i.e., gain coefficients k_i are larger than in the first experiment. Figs. 17 and 18 show number of users and revenue, respectively, for the algorithm that does use CAC mechanism. In fact, k_i is now so large that the algorithm accepts in practice every call, and this is seen from Fig. 17, i.e., $N_i(t)$ is approximately the same as stated by Little's formula. Fig. 18 tells us that the revenue almost achieves the upper limit of Theorem 2 due to the dominating role of $\sum_i N_i k_i$ in Theorem 3.

V. DISCUSSION AND FUTURE WORK

A. Properties of the Closed Form Scheduling Algorithm

Here, we present conclusions from our approach as well as experiments. Also some future topics are discussed. The conclusions are as follows:

- We have analytically shown that in the case of linear pricing scenario, revenue has unique maximum in the weight interval $w_i \in (0, 1]$. Proof was based on the closed form solution of the weights as well as concavity of the revenue function in the interval $w_i \in (0, 1]$.
- The proposed weight updating algorithm is computationally inexpensive in our scope of study.
- Experiments clearly justify the performance of the algorithm. For example, theorems for upper bounds hold, and revenue curves are positive.
- Some of the statistical and deterministic algorithms presented in the literature assume quite strict *a priori* information about parameters or statistical behavior such as call densities, durations or distributions. However, such methods usually are - in addition to being computationally complex - not robust against erroneous assumptions or estimates. On the contrary, our algorithm is deterministic and non-parametric, i.e., it uses only the information about the number of customers, and thus we believe that the robustness makes it a competitive candi-

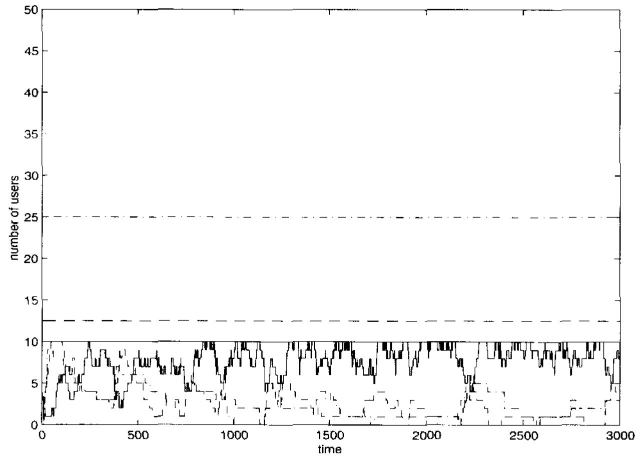


Fig. 15. Second experiment with call admission control. Number of users as a function of time. Horizontal axis: time. Vertical axis: number of users. Solid, dashed, and dash-dotted curves correspond to gold, silver, and bronze class, respectively. Lines are upper bounds for buffer sizes.

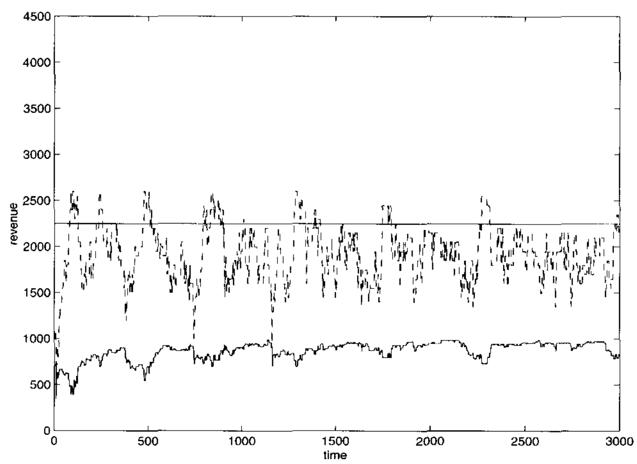


Fig. 16. Second experiment with call admission control. Revenue as a function of time. Horizontal axis: time. Vertical axis: revenue. Lowest curve: realized revenue; middle curve (dashed): upper bound $\sum_i N_i k_i$; solid line: $0.25 \sum_i k_i^2/r_i$.

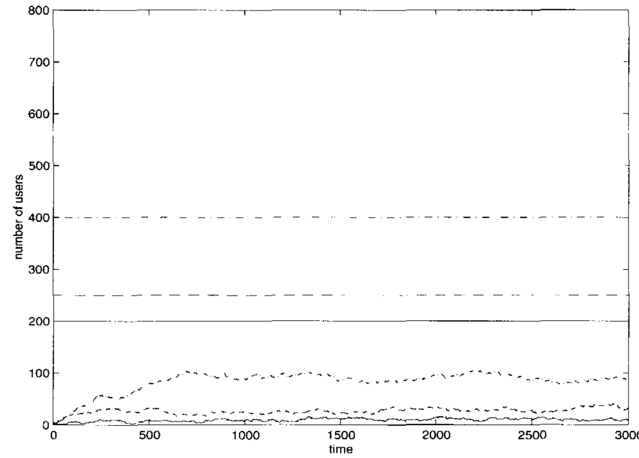


Fig. 17. Third experiment with call admission control. Number of users as a function of time. Horizontal axis: time. Vertical axis: number of users. Solid, dashed, and dash-dotted curves correspond to gold, silver, and bronze class, respectively. Lines are upper bounds for buffer sizes.

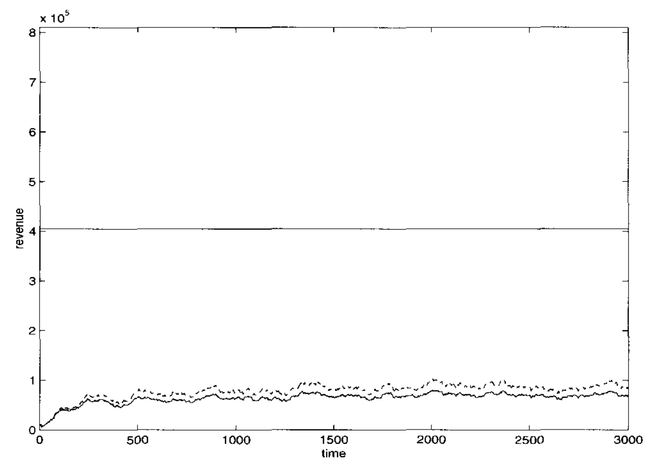


Fig. 18. Third experiment with call admission control. Revenue as a function of time. Horizontal axis: time. Vertical axis: revenue. Lowest curve: realized revenue; middle curve (dashed): upper bound $\sum_i N_i k_i$; solid line: $0.25 \sum_i k_i^2/r_i$.

date in practical environments.

- Also, CAC mechanism can be used in the context of the algorithm. It is based on the hypothesis testing, and is computationally quite simple.
- Here, we investigated only the single node case. Multinode case is a much more challenging problem.
- The algorithm used the same packet sizes. However, it is quite straightforward to develop the version, which can handle different packet sizes.

General conclusion is that the linear pricing scenario is quite simple, perhaps tempting, and practical. However, we believe that more practical pricing scheme should be based on *piecewise* linear model. Studies are made in that direction.. Especially flat pricing scenario is an interesting topic of study.

B. Extensions of the Scenario

As we mentioned earlier, in the more general case, every traffic flow has different routes, and therefore different delay constants c_{ij}. For reducing the number of parameters, one can approximate

$$F = \sum_{i=1}^{m} \sum_{j=1}^{n} N_{ij} r_i (N_{ij} d_0 / w_i + c_{ij}), \tag{47}$$

where N_{ij} is the number of customers having (approximately) constant delay c_{ij}. However, in the most accurate criterion, there are different weights for different user numbers N_{ij}, and revenue criterion has the form

$$F = \sum_{i=1}^{m} \sum_{j=1}^{n} N_{ij} r_i (N_{ij} d_0 / w_{ij} + c_{ij}). \tag{48}$$

But we see that this is quite a trivial extension, because (48) can be re-organized to the form

$$F = \sum_{i=1}^{mn} \tilde{N}_i \tilde{r}_i (\tilde{N}_i d_0 / w_i + \tilde{c}_i), \tag{49}$$

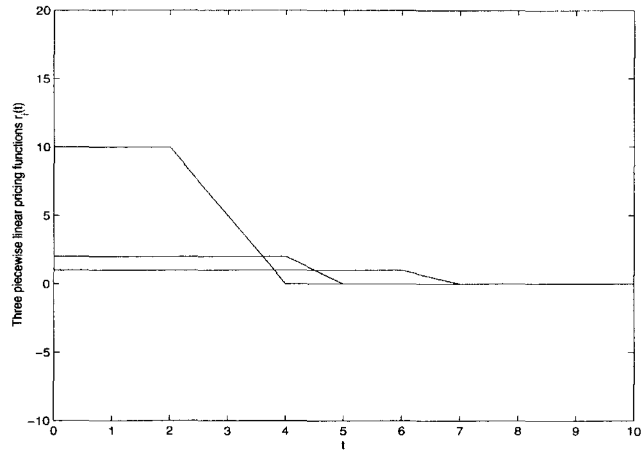


Fig. 19. Three linear pricing functions, where minimum price is $r_i(d) = 0$. Horizontal axis: delay; vertical axis: price.

$$r_1 = r_2 = \cdots = r_n = \tilde{r}_1, \tag{50}$$

$$r_{n+1} = r_{n+2} = \cdots = r_{2n} = \tilde{r}_2, \tag{51}$$

$$\cdots,$$

$$c_{ij} = \tilde{c}_i, \quad i = 1, \cdots, m, \tag{52}$$

$$\tilde{N}_i = N_{(i-1)n+j}. \tag{53}$$

Thus, only computational complexity increases by the factor n to $O(mn)$ multiplications and additions, where n is number of different quantized delay values. Because delays are continuous, an approximation

$$\tilde{c}_1 \in [0, \Delta), \tag{54}$$

$$\tilde{c}_2 \in [\Delta, 2\Delta), \tag{55}$$

$$\cdots,$$

$$\tilde{c}_n \in [(n-1)\Delta, n\Delta), \tag{56}$$

must be made. Here time Δ (seconds) depends on the application.

To see that the basic closed form solution (10) does not need essential modifications, when the delays c_i are included in the revenue F, we write F in the linear scenario as follows:

$$r_i(d_i) = r_i(N_i/w_i + c_i), \tag{57}$$

$$\begin{aligned} F &= \sum_{i=1} N_i[-r_i(N_i/w_i + c_i) + k_i] \\ &= -\sum_{i=1}^{m} \frac{r_i N_i^2}{w_i} + \sum_{i=1}^{m} N_i(k_i - r_i c_i) \\ &= -\sum_{i=1}^{m} \frac{r_i N_i^2}{w_i} + \sum_{i=1}^{m} N_i \tilde{k}_i, \end{aligned} \tag{58}$$

where change of variable $\tilde{k}_i = k_i - r_i c_i$ has been made. This is essentially same as (9). However, the formula is important, because it gives us constraint

$$k_i > r_i c_{i,max}, \tag{59}$$

to avoid negative revenue. Most straightforward solution for avoiding $F < 0$ is to use piecewise linear model. One candidate should be to continuously use (10) updating rule (which is based on the assumption that negative revenue is possible), while minimal value to the pricing functions are selected to be $r_i(d) = 0$, i.e., service provider does not need to pay money to the customer, but give free service when delay tends to be too large. Fig. 19 illustrates the corresponding pricing functions.

VI. CONCLUSIONS

In this paper, we introduced a closed form scheduling algorithm, which was derived from revenue target function by Lagrangian optimization approach. The experiments demonstrated the revenue maximization ability of the algorithm, while still allocating delays in a fair way.

In the future work, multinode case is investigated. It is important to develop such a distributed approximation, which does not suffer the curse of dimensionality and computational complexity of the optimal global approach. Also different kind of pricing models will be studied. We have also started to work with Linux routers, and the goal of the task is to implement presented algorithm to a real router environment.

REFERENCES

[1] J. K. MacKie-Mason and H. R. Varian, *Pricing the Internet, in Public Access to the Internet*, NJ: Prentice-Hall, 1994.

[2] F. P. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecommun.*, vol. 8, pp. 33–37, 1997.

[3] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *Oper. Res. Soc.*, vol. 49, pp. 237–252, 1998.

[4] R. J. Gibbens and F. P. Kelly, "Resource pricing and the evolution of congestion control," *Automatica*, vol. 35, no. 12, pp. 1969–1985, 1999.

[5] S. Kunniyur and R. Srikant, "End-to-end congestion control: Utility functions, random losses, and ECN marks," in *Proc. INFOCOM 2000*, 2000, pp. 1323–1332.

[6] R. J. La and V. Anantharam, "Utility-based rate control in the internet for elastic traffic," *IEEE/ACM Trans. Networking*, vol. 10, pp. 272–286, Apr. 2002.

[7] S. H. Low, "Equilibrium bandwidth and buffer allocations for elastic traffics," *IEEE/ACM Trans. Networking*, vol. 8, pp. 373–383, June 2000.

[8] I. Ch. Paschalidis and J. N. Tsitsiklis, "Congestion-dependent pricing of network services," *IEEE/ACM Trans. Networking*, vol. 8, pp. 171–184, Apr. 2000.

[9] I. Ch. Paschalidis and Yong Liu, "Pricing in multiservice loss networks: Static pricing, asymptotic optimality, and demand substitution effects," *IEEE/ACM Trans. Networking*, vol. 10, pp. 425–438, June 2002.

[10] F. P. Kelly, *Notes on Effective Bandwidths, in Stochastic Networks: Theory and Applications*, Oxford Univ. Press.

[11] I. Ch. Paschalidis, "Class-specific quality of service guarantees in multimedia communication networks," *Automatica*, vol. 35, no. 12, pp. 1951–1968, 1999.

[12] F. P. Kelly, "On tariffs, policing, and admission control for multiservice networks," *Oper. Res. Lett.*, vol. 15, pp. 1–9, 1994.

[13] C. Courcoubetis, F. P. Kelly, and R. Weber, "Measurement-based usage charges in communications networks," *Oper. Res.*, vol. 48, no. 4, pp. 535–548, 2000.

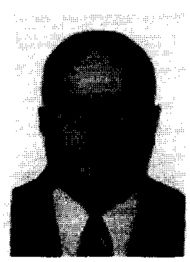
[14] A. Orda, R. Rom, and N. Shimkin, "Competitive routing in multiuser communication networks," *IEEE/ACM Trans. Networking*, vol. 1, no. 5, pp. 510–521, Oct. 1993.

[15] Y. Korilis, A. Lazar, and A. Orda, "Architecting noncooperative networks," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1241–251, 1995.

[16] Y. A. Korilis, A. A. Lazar, and A. Orda, "Achieving network optima using Stackelberg routing strategies," *IEEE/ACM Trans. Networking, vol. 5*, pp. 161–173, Feb. 1997.

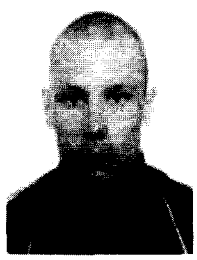
[17] R. J. La and V. Anantharam, "Optimal routing control: A game theoretic approach," in *Proc. Decision and Control*, San Diego, CA, Dec. 1997.

[18] E. Altman, T. Basar, and R. Srikant, "Nash equilibria for combined flow control and routing in networks: Asymptotic behavior for a large number of users," *IEEE Trans. Automat. Contr.*, vol. 47, pp. 917–930, June 2002.
[19] E. Altman *et al.*, "Competitive routing in networks with polynomial costs," *IEEE Trans. Automat. Contr.*, vol. 47, pp. 92–96, Jan. 2002.
[20] T. Basar and R. Srikant, "Revenue-maximizing pricing and capacity expansion in a many-users regime," in *Proc. INFOCOM 2002*, vol. 1, 2002, pp. 294–301.
[21] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The single node case," *IEEE/ACM Trans. Networking*, vol. 1, no. 3, pp. 344–357, June 1993.
[22] J. C. R. Bennett and H. Zhang, "WF2Q : Worst-case fair weighted fair queueing," in *Proc. INFOCOM'96*, 1996.
[23] A. Demers, S. Keshav, and S. Shenker, "Analysis and Simulation of a fair queueing algorithm," in *Proc. SIGCOM'89*, 1989, pp. 1–12.
[24] N. Matsufuru and R. Aibara, "Efficient fair queueing for ATM network using uniform round robin," in *Proc. INFOCOM'99*, 1999, pp. 389–397.
[25] H. M. Chaskar and U. Madhow, "Fair scheduling with tunable latency: A round robin approach," in *Proc. GLOBECOM'99*, 1999, pp. 1328–1333.
[26] J.-S. Kim and D. C. Lee, "Weighted round robin packet scheduler using relative service share," in *Proc. MILCOM 2001*, vol. 2, 2001, pp. 988–992.
[27] J. Joutsensalo and T. Hämäläinen, "QoS-aware adaptive pricing for network Services," in *Proc. GLOBECOM 2001*, vol. 4, 2001, pp. 2455–2459.
[28] J. Joutsensalo and T. Hämäläinen, "Optimal link allocation and revenue maximization," *J. Commun. Networks*, vol. 4, no. 2, pp. 136–147, June 2002.
[29] T. Hämäläinen and J. Joutsensalo, "Link allocation and revenue optimization for future networks," in *Proc. GLOBECOM 2002*, Taiwan, Nov. 2002.
[30] J. Joutsensalo, *et al.*, "Adaptive weighted fair scheduling method for channel allocation," in *Proc. ICC 2003*, Anchorage, Alaska, 2003.

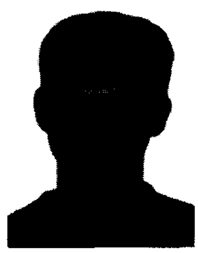


J. Joutsensalo was born in Kiukainen, Finland, in July 1966. He received diploma engineer, licentiate of technology, and doctor of technology degrees from Helsinki University of Technology, Espoo, Finland, in 1992, 1993, and 1994, respectively. Currently, he is Professor of Telecommunications at the University of Jyväskylä. His research interests include signal processing for telecommunications, as well as data networks and pricing.

T. Hämäläinen received the B.Sc in automation engineering from the Jyväskylä Institute of Technology in Finland on 1991 and the M.Sc and the Ph.D. degrees in telecommunication from Tampere University of technology and University of Jyväskylä, Finland in 1996 and 2002, respectively. His Ph.D. works studied Quality of service and pricing issues in broadband networks. His current research interests include traffic engineering and QoS in wired and wireless networks.



M. Pääkkönen received the M.Sc degree in telecommunication from University of Jyväskylä, Finland, in 2001. His current research interests lie in the areas of traffic engineering, QoS and pricing issues in broadband networks.



A. V. Sayenko has obtained the B.Sc degree from the Kharkov State University of RadioElectronics (Ukraine) in 2001. In 2002, he has obtained the M.Sc degree from the University of Jyväskylä (Finland), in which he works now on towards the Ph.D. His scientific interests include Quality-of-Service, network and policy-based management technologies.