Generic model for revenue maximization

Antti Niemi, Jyrki Joutsensalo, Timo Hämäläinen
University of Jyv¨askyl¨a
Faculty of Information Technology
Department of Mathematical Information Technology
P.O. Box 35, FIN-40351 Jyv¨askyl¨a, FINLAND
[1] $JyvaskylaPolytechnic$

ABSTRACT

This paper introduces a model that can be used to share link capacity among customers under different kind of traffic conditions. This model is suitable for different kind of networks like the 4G networks (fast wireless access to wired network) to support connections of given duration that requires a certain quality of service. We study different types of network traffic mixed in a same communication link. A single link is considered as a bottleneck and the goal is to find customer traffic profiles that maximizes the revenue of the link. Presented allocation system accepts every calls and there is not absolute blocking, but the offered data rate/user depends on the network load. Data arrival rate depends on the current link utilization, user's payment (selected CoS class) and delay. The arrival rate is (i) increasing with respect to the offered data rate, (ii) decreasing with respect to the price, (iii) decreasing with respect to the network load, and (iv) decreasing with respect to the delay. As an example, explicit formula obeying these conditions is given and analyzed.

Keywords: Pricing, quality of service, performance.

I. Introduction

We consider a link allocation scheme that can be used to tune link utilization and per connection QoS as high as possible. Our model can be used with different kind of networks e.g broadband, wireless and IP -networks. At the packet level the link allocation scheme can be seen as illustrated in Fig 1.

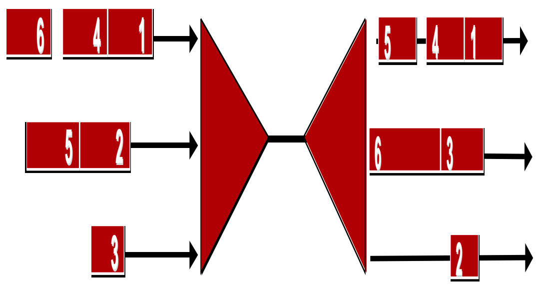

Fig. 1. An example of single link allocation

The approach studied in this paper has similar features as used in [3], [7], [8], [10]. These studies mark each packet entering a switch during a congestion. Marks reflect the fact that these packets imposed a cost in terms of delay or loss on some other packet. End users are informed of whether their packet was marked when acknowledgements flow back to them and are then free to decide how to adapt their transmission rates.

Our model does not require the core network to maintain per flow state. The task of associating prices with end users is pushed to the network edges where it is easier to deal with. We propose a mechanism for users to respond to price information which is motivated by the assumption that they are trying to maximize their individual utilities. A related scheme has studied in [9] in the limit where the number of users increases to infinity. In this asymptotic regime, the price turns out to be constant and so the model does not capture dynamics. In this paper, we continue the work presented in [4], [5]. In [5] we considered data rate allocation, while in this paper we consider also user's reaction to the offered delay. The new features in this paper are:

- Arrival data rate depends on the offered data rate, delay, and current link utilization as well as price.
- The smaller the total demand is, more popular are cheap CoS classes.
- The smaller the total demand is, smaller is the total revenue.
- Stable state approximation of the arrival rate is studied.

The rest of the paper is organized as follows. In Section 2 we present link allocation scenario, and in Section 3 the arrival rate model is shown. In Section 4 stable state approximation is presented. Section 5 contains simulations and results, and the final section concludes the study.

II. Allocation scenario

Let A be customer and B be the operator or manager network administrator. *Price* x streams from A to B. The price paid by A is normalized to be

$$x \in (0,1) \tag{1}$$

User profile function $n(x,t)$ depends on both price x and time t, and it has the property

$$n(x,t) > 0 \tag{2}$$

for all defined values of x. $n(x_0,t_0)$ tells us, how many users pay x_0 money units at time instant t_0. Positivity of the function means that A pays money to B, but not the opposite price stream is possible in this study. $n(x,t)$ is defined for continuous $x \in (0,1)$, but in practical application, $n(x,t)$ is sampled and vectorized.

Definition of an *data rate function* - which belongs to the class of *utility functions* - $u(x,t)$ is the following. When the price x streams from A to B, the data rate $u(x,t)$ streams between the corresponding users A via B. Data rate $u(x,t)$ depends on x, and it is a strictly increasing function of x. Generally speaking,the utility is a service rate, or Quality of Service (QoS). In telecommunications applications, utility may include e.g. data rate, Bit Error Rate (BER), delay, or blocking probability. Here we assume that $u(x,t)$ is scalar with respect to scalar x and t.

B tries to keep all the capacity of the channel in use, so that the customers A are as satisfied as possible. By doing this B also maximizes his revenue. *Channel capacity* C is the maximal number of information that the entire network B can transfer without errors. This is the basic definition given by classical Shannon information theory [2]. In practice, suboptimal estimate of C is obtained by observing the data flow history in the network. In this paper, we do not take a stance on the strict estimation procedure of the channel capacity.

In our scenario the operator offers the data rate as follows:

$$u(x,t) = \gamma(t) f(x, \alpha_1, \ldots, \alpha_n) \tag{3}$$

Here the basic function $f(x) = f(x, \alpha_1, \ldots, \alpha_n)$ has the following properties:

- $f(x)$ strictly increases with respect to the price x.
- $f(0) = 0$, i.e. no utility is allocated when no money is payed.
- Without loss of generality, the maximum value is assumed to be $f(1) = 1$.
- The parameters α_i are optimized for maximizing the revenue.

The are infinitely the basic functions satisfying the above conditions. One possible basic function - which we have selected as a test case - has the form

$$f(x, \alpha) = \frac{1 - e^{-\alpha x}}{1 - e^{-\alpha}} \tag{4}$$

where α is a parameter to be optimized for maximizing the revenue. Eq. (4) obeys $f(0) = 0$ and $f(1) = 1$. We also assume that all connection request are accepted, and the overall data rate will be kept as high as possible. In other words, the equality

$$\int_0^1 u(x,t) n(x,t) dx = 1 \tag{5}$$

is tried to kept at all the time t. Here the capacity C is normalized to be $C = 1$. Then it follows from Eqs. (3) and (5) that

$$\gamma(t) = \frac{1}{\int_0^1 f(x) n(x,t) dx} \tag{6}$$

and the customer, who pays the price x at the time t, gets the data rate

$$u(x,t) = \frac{f(x)}{\int_0^1 f(y) n(y,t) dy} \tag{7}$$

Notice that the given data rate dynamically varies with respect to the time t due to existence of the user profile function $n(y,t)$ in the denominator of Eq. (7). When $n(x,t)$ - i.e. the number of the users - is small, the integral in the denominator of Eq. (7) is small, and as a consequence, the offered data rate/user class is high. On the other hand, in the highly loaded systems, the offered data rate/user class is low.

Operator offers also a delay profile function $g(x,\beta)$, where the inverse $g(x,\beta)$ of the delay is measured by $1/$seconds, and is strictly increasing with respect to the price x. The customer knows the delay utility function $1/g$ (seconds).

III. Arrival Rate Model

In [4], we presented the arrival rate model which depended only on one QoS parameter, namely offered data rate, via functions $f(x,\alpha)$ and $\gamma(t)$. In this paper, the arrival rate $\lambda(x,t)$ is a function of several QoS parameters such as data rate, delay, or BER. The general model for the $\lambda(x,t)$ in this study is

$$\lambda = \lambda[x, t, \alpha_1, \ldots, \alpha_n, \gamma_1(t), \ldots, \gamma_m(t)] \tag{8}$$

where

- α_i are called *hidden QoS parameters* to be optimized. That is, the customers do not observe those parameters.
- $\gamma_i(t)$ are called *observed time-varying QoS parameters*. That is, users observe those parameters, for example $\gamma(t)$ (Eq. 6). More precisely, the customer knows the offered time-variant and time-invariant utility functions which contain the information about $\gamma(t)$. For example, time-variant $u(x,t)$ and time-invariant $f(x,\alpha)$ are known, and so $\gamma(t) = u(x,t)/f(x,\alpha)$ can be observed if necessary. $f(x,\alpha)$ is available in tabular or explicit functional form.

One special case of the general form is

$$\lambda(x,t) = \frac{p(x, \alpha_1, \ldots, \alpha_n)}{1 + \int_0^1 e(y, \beta_1, \ldots, \beta_m) n(y,t) dy} \tag{9}$$

where $p > 0$ and $e > 0$ are some, perhaps very complicated, functions of the offered QoS parameters. Positivity of p and e implies positivity of λ.

In this paper, we study the simple case where arrival rate depends on the following issues:
(i) The larger is offered basic function $f(x,\alpha)$, the larger is the call density.
(ii) The larger is price x, the smaller is call density.
(iii) The larger is load, the smaller is call density.
(iv) The larger is the inverse $g(x,\beta)$ of the delay, the larger is the call density.

One special case we selected [4], [5]

$$p(x, \alpha, \xi) = k f(x, \alpha) h(x, \xi) \tag{10}$$

$$e(x, \alpha) = f(x, \alpha) \tag{11}$$

and $\lambda(x, t)$ could then be as in the following equation:

In the model obeying better the condition (the smaller total demand is, more popular are cheap QoS classes), the arrival rate $\lambda(x, t)$ is as follows:

$$\lambda(x, t) = \frac{k\xi(t)^z f(x, \alpha)h(x, \xi)}{1 + 1/\gamma(t)} \tag{12}$$

where $z \geq 1$. Here $\xi(t)^z k$ can be thought as a new scalar coefficient $k(t) = \xi^z(t)k$, and the stable state model remains essentially the same as in Eq. (24). The difference is that the revenue must be evaluated for different values of $\xi(t)$.

The model of $\lambda(x, t)$ is decomposed into three sub-functions. In the model (12), $f(x, \alpha)$ represents the feature (i), i.e. the arrival rate increases with respect to the offered data rate. Typically $h(x, \xi)$, which represents the feature (ii), depends on the distribution of the richness or the willingness of pay of the users, and therefore the general properties of $h(x) = h(x, \xi)$ are:

- $h(x)$ strictly decreases with respect to the price x.
- Without loss of generality, $h(0) = 1$, i.e. the willingness of pay is maximum.
- Without loss of generality, the minimum is at $h(1) = 0$.
- The parameter ξ controls the behavior of the curve.

In this study, we have for simplicity the exponentially decreasing convex form

$$h(x, \xi) = \frac{e^{-\xi x} - e^{-\xi}}{1 - e^{-\xi}} \tag{13}$$

where $\xi > 1$. For large ξ, the function $h(x, \xi)$ is abrupt. It is natural that in our scenario the denominator of (12) consists of the feedback term $\int_0^1 f(y, \alpha)n(y, t)dy$, since that term includes the information about the offered data rate per user class, as shown in Eq. (7). That integral represent the feature (iii). Notice that the arrival rate can also be written in the form

$$\lambda(x, t) = \frac{kf(x, \alpha)h(x, \xi)}{1 + 1/\gamma(t)} \tag{14}$$

which shows the dependence of the arrival rate with respect to the price, time, and the given data rate per QoS class. When no users exist in the network, maximal arrival rate

$$\lambda_{\max}(x, t) = kf(x, \alpha)h(x, \xi) \tag{15}$$

is achieved.

In this work, the delay is QoS parameter, too, and the functional form of p in Eq. 9 is

$$p = p(x, f(x, \alpha), g(x, \beta))h(x, \xi) \tag{16}$$

where p is strictly increasing with respect to f and g. One example of a functional dependence is

$$p(x) = kf(x, \alpha)g(x, \beta)h(x, \xi) \tag{17}$$

where p has been separated to three sub-functions. In this more general optimization problem, α defines the offered minimum data rate, while β defines the offered maximum delay.

In the experiments, we use for simplicity the function

$$p(x) = kf(x, \alpha)^z h(x, \xi) \tag{18}$$

where the user's reaction - and thus the dependence of the arrival rate - to the offered QoS parameters, namely data rate and delay, has been embedded to the one function, $f(x, \alpha)^z$. Thus here it is assumed that

$$g(x, \alpha) = f(x, \alpha)^{z-1} \tag{19}$$

so that Eqs. (17) and (18) equal. The offered delay $1/f(x, \alpha)^{z-1}$ as a function of the price x is plotted in Fig. 7 by using the value $z = 1.5$.

IV. Stable State Analysis

In [5], we derived the case where the system is in steady state or near it, and $\lambda(x, t)$ changes sufficiently slowly according to the arrival model (Eqs. 9,11). We can use the Little's formula [1] in the theoretical analysis as follows:

$$E(n(x, t)) = \frac{E(\lambda(x, t))}{\mu(x)} = \frac{\lambda(x)}{\mu(x)} \tag{20}$$

The above assumption is valid for large number of different types of stochastic processes, when λ is the arrival rate and $1/\mu$ is the expected value of the connection time.

The expected revenue per time of the system is

$$revenue = \int_0^1 xE[n(x, t)]dx = \lim_{T\to\infty} \frac{1}{T} \int_0^T \int_0^1 xn(x, t)dxdt \tag{21}$$

and by using Little's formula, it is

$$revenue = \int_0^1 x\frac{\lambda(x)}{\mu(x)}dx \tag{22}$$

Thus, if $p(x)$ and $\mu(x)$ are known, we can numerically find the optimal α that maximizes revenue (22) by using formulae below.

$$revenue = \int_0^1 x\frac{\lambda(x, \alpha)}{\mu(x)}dx \tag{23}$$

where

$$\lambda(x, \alpha) = \frac{-1 + \sqrt{1 + 4q(\alpha)}}{2q(\alpha)}p(x) \tag{24}$$

$$q(\alpha) = \int_0^1 f(x, \alpha))\frac{p(x)}{\mu(x)}dx \tag{25}$$

V. Simulations and Results

In the simulations, we compare the revenue (21) given by simulated network traffic data with that revenue (23) given by the analytical model. Our goal is to examine which kind of shape of the capacity function $f(x,\alpha)$ in Eq. (4) could be optimal for maximizing the revenue under two arrival rate scenarios: in the more general case (9), and in the special case (12). Our other goal is to prove that the analytical models in the special case (23)-(25) as well as in the more general case (23), (24), (25) hold well enough. If these models hold, then one can easily construct the revenue curves corresponding to the different data traffic scenarios and utility function families, and therefore obtain estimates for optimal data rate allocation strategies.

In the first experiment, we study the behavior of the arrival rate and the revenue maximization by varying ξ. Figure 2 shows the variation of ξ as a function of time. Here the scaling parameters are $k_1 = 2$ and $k_2 = 0.1$. It is seen that the fluctuation of ξ is quite random due to the feedback $\int_0^1 f(x,\alpha)n(x,t)dx$. Therefore, the behavior of the arrival rate $\lambda(x,t)$, as shown in Fig. 3, is quite non–stationary. However, the revenue curve shown in Figure 4 is quite similar than the curves obtained in the earlier simulations by fixed ξ. Figs. 5 and 6 depict the revenue using a stable state model and a simulation model as a function of α with values of $\xi = 1, 2, \ldots, 10$. Curves in these Figs. match quite well showing that the steady-state model holds very well in this special case, too.

Next, we perform the simulation where z takes the values $1.01, 1.02, \ldots, 2$. The number of different user classes was 99. Figure 8 shows the family of the curves $p(x) = kf(x,\alpha)^z h(x,\xi)$ with different values of z, and $\xi = 2$. The optimal α maximizing the revenue (Eq. 23) is obtained by evaluating Eqs. (24) and (25), and it is represented as a function of z in Fig. 9. Because optimal $\alpha > 0$ for all $z \in [0.01, 2]$, the optimal utility functions $f(x,\alpha)$ are concave. This is plausible result, because of the shape of $h(x,\xi)$ will lead to the fact the rich users are ready to pay much more money while obtaining only small differential increase in the utility, say data rate and delay. This is common results known in other service cases, e.g. in the airplanes first, business and economy classes exist.

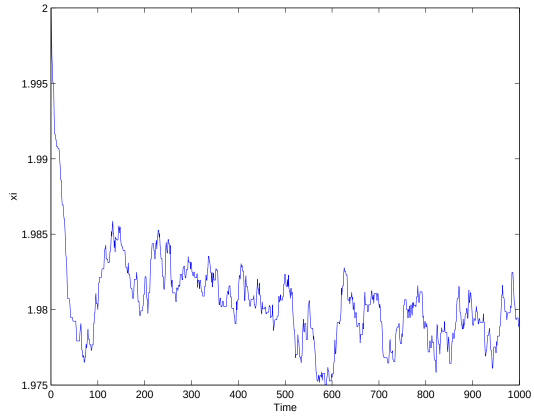

Fig. 2. Variation of ξ as a function of time.

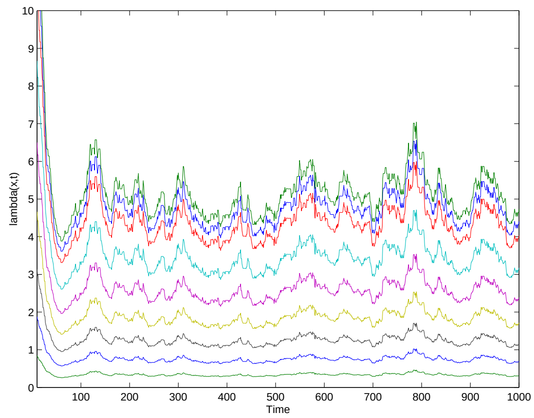

Fig. 3. Behavior of the arrival rate $\lambda(x,t)$.

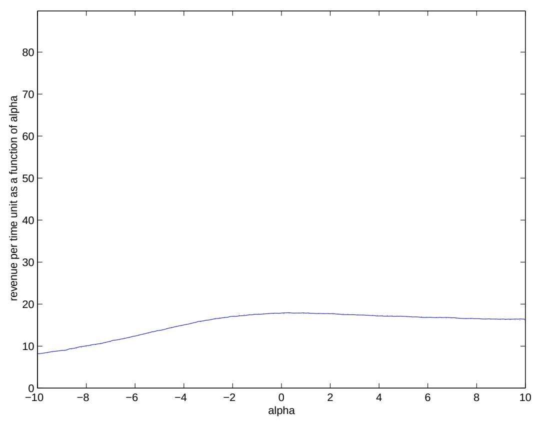

Fig. 4. Revenue as a function of α.

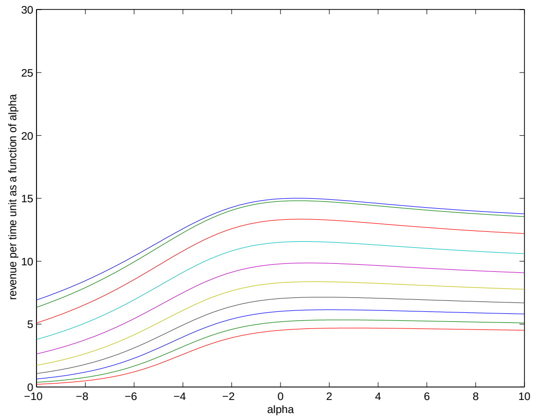

Fig. 5. Revenue using a stable state model as a function of α with values of $\xi = 1, 2, \ldots, 10$.

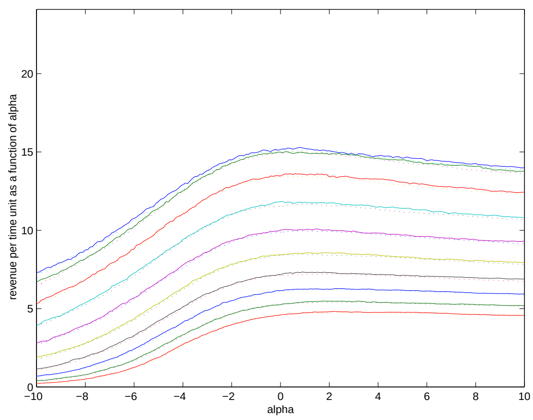

Fig. 6. Revenue using a simulation model as a function of α with values of $\xi = 1, 2, \ldots, 10$.

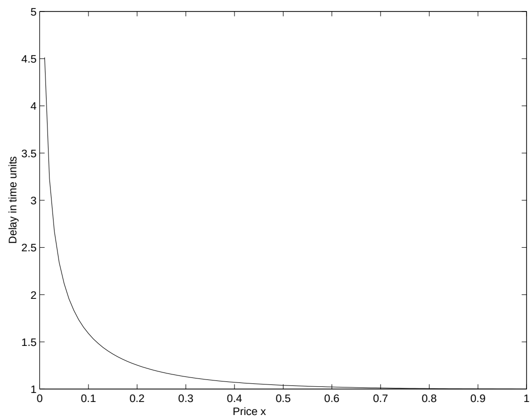

Fig. 7. Delay in time units as a function of price x, when $\alpha = 5$ and $z = 1.5$

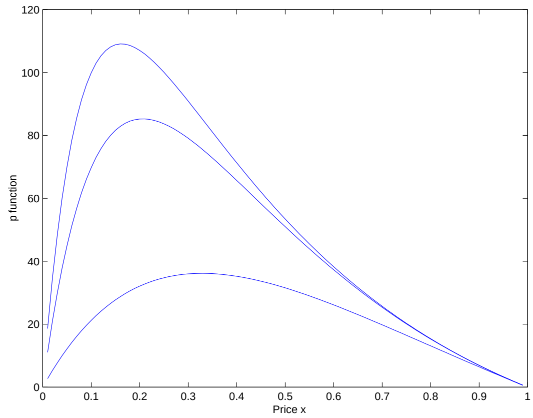

Fig. 8. Three different $p(x)$ functions with parameters $1, 1.5, 2$

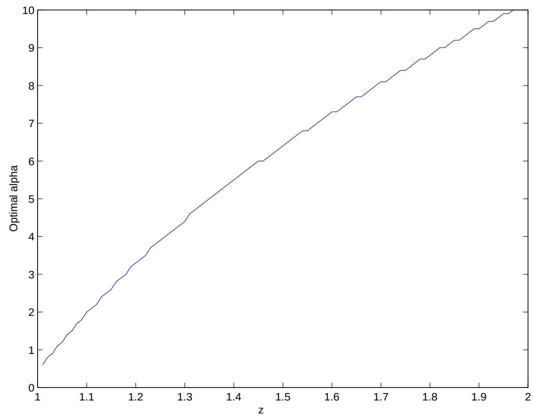

Fig. 9. Optimal α as a function of z, when the arrival rate depends on the data rate and delay.

VI. Conclusions

In this paper, we generalize our channel allocation model further by assuming that the demand varies as a function of time. We add the following assumption:

(i) The smaller total demand is, more popular are cheap QoS classes.

(ii) The smaller the total demand is, smaller is the total revenue.

As an example, explicit formula obeying these conditions was given and analyzed. Results showed that our steady-state model holds very well in the above special case, too.

Future work...

References

[1] J. Banks, J. S. Carson II and B. L. Nelson, ”Discrete-Event System Simulation,” Prentice-Hall, Inc. 1996.

[2] T.M. Cover and J.A. Thomas, ”Elements of Information Theory”, Wiley 1991.

[3] R. J. Gibbens and F. P. Kelly, ”Resource pricing and the evolution of congestion control”, Automatica, vol. 35, pp. 1969–1985, 1999.

[4] T. H¨am¨al¨ainen and J. Joutsensalo, ”QoS-Aware Adaptive Pricing for Network Services”. Proceedings of IEEE Globecom 2001, November 2001, USA.

[5] J. Joutsensalo and T. Hämäläinen, "Optimal Link Allocation and Revenue Maximization". Journal of Communications and Networks, in press.
[6] F. P. Kelly, P. B. Key, and S. Zachary, "Distributed admission control", IEEE JSAC, vol. 18, pp. 2617–2628, 2000.
[7] F. P. Kelly, "Charging and rate control for elastic traffic", European Transaction on Telecommunications, vol. 8, pp. 33–37, 1997.
[8] F. P. Kelly, A. Maulloo and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability", Journal of the Operational Research Society, vol. 49, pp. 237-252, 1998.
[9] P. Key and L. Massoulie, "User policies in a network implementing congestion pricing", MIT-Tufts Internet Service Quality Economics Workshop, Boston, 1999.
[10] I. Ch. Paschalidis and J. N. Tsitsiklis, "Congestion-Dependent Pricing of Network Services", IEEE/ACM Transactions on Networking, vol. 8, pp. 171-184, 2000.