FAST CLOSED FORM APPROXIMATION FOR DYNAMIC NETWORK RESOURCE ALLOCATION

[*]Jyrki JOUTSENSALO, [**]Isto KANNISTO, [***]Jarmo SILTANEN, [*]Timo HÄMÄLÄINEN
[*]Department of Mathematical Information Technology, University of Jyväskylä,
40014 University of Jyväskylä, Finland, Telephone: +358 14 2603292,
E-mails: jyrkij@mit.jyu.fi, timoh@cc.jyu.fi
[**]TeliaSonera, Jyväskylä, Finland, E-mail: isto.kannisto@teliasonera.com
[***]Jyväskylä Polytechnic, Finland, E-mail: jarmo.siltanen@jypoly.fi

SUMMARY

We consider dynamic delay guarantee and bandwidth allocation in communications networks. Our scenario includes linear pricing scheme for both Quality of Service parameters. The goal is (i) to maximize the revenue and (ii) guarantee fair resource allocation for connections. On the contrary to the traditional Lagrangian approach, we approach the problem by modified one, where the sum of the weights of the scheduler acts as the penalty term. This modified approach yields closed form approximate algorithm for updating the scheduler weights, being very fast and real-time implementable. We compare the algorithm with the brute-force method, which optimizes weights in the large grid - optimal brute-force method has exponential complexity. The revenue obtained by the closed form method is about 99.9 % of the optimal, computationally expensive approach, thus being tempting both from the point of view of the service provider and the customers. NS-2 simulator is used in the experiments.

Keywords: *Quality of Service, Network Resource Allocation, pricing models, computer networks*

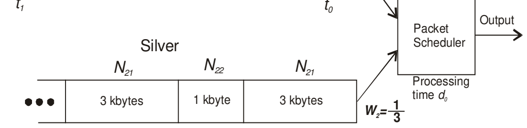
1. INTRODUCTION

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 [24]. 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 [14], [15]. Packet-based pricing schemes (e.g. [8], [20]) 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 [21]. 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 [23]. Pricing and link allocation for real-time traffic that requires strict QoS guarantees is studied e.g. in [27], [28]. 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., [16] for a survey of effective bandwidth characterizations and [26] for similar notions in the multiclass case). In this setting, [17], [6] 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 realtime mechanism is needed to communicate tariffs to the users.

There is also several research work done with the gametheoretic models of routing and flow control in communication networks (e.g. [30], [18], [19], [22], [1], [2]). 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 [18]. 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 [22]. A revenue-maximizing pricing scheme for the service provider is presented in [3]. Thus, a noncooperative (Nash) flow control game is played by the users (followers) in a Stackelberg gamewhere the goal of the leader is to set a price to maximize revenue.

Two well-known scheduling algorithms are the packet-bypacket generalized processor sharing (PGPS) ([29]) and the worstcase fair weighted fair queueing (WF2Q) ([4]). The WF2Q 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 [29], which is equivalent to the weighted fair queuing (WFQ) [7]. As outlined in [4], 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., WF2Q) was proposed in [4]. The queueing disciplines such as PGPS and WF2Q 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 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 [25] and [5], the uniform round robin (URR) discipline and the WF2Q interleaved WRR discipline emulate the WF2Q 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 [10]. They present a scheduling algorithm for fixed length packets that do not emulate the WF2Q. 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, [31], [32], to take into account queuing scheduling issues by introducing dynamic weight tracking algorithm in the scheduler. QoS and revenue aware scheduling algorithm is investigated. It is derived from optimization problem, that resembles Lagrangian constrained approach, and approximate optimal closed form solution is presented, when QoS parameters are delay and bandwidth.



The rest of the paper is organized as follows. In Section 2, used pricing scenario is presented and generally defined. Closed form scheduling algorithm is derived in Section 3; in addition, Call Admission Control (CAC) mechanism as well as some upper bounds are presented in this section. Section 4 contains experimental part justifying theorems. Discussions are made in Section 5, and final section contains conclusions of the work.

2. DELAY AND BANDWIDTH MODELS

In this section, we formulate expressions for delays (seconds) and bandwidth (bit rate) of the data traffic. Consider the packet scheduler for two service classes. There are now two service classes. Gold class customers pay most of money while getting best service, and silver class pay less of money. Bronze class customers pay least of money while getting worst service.

Parameter Δt_i denotes time which passes when data is transferred through the queue i to the output in the switch, when $w_i = 1$. If the queue is almost empty, delay is small, and when the buffer is full, it is large. Variable w_i is the weight allocated for class i. Constraint for weights w_i is

$$\sum_{i=1}^{m} w_i = 1, \quad w_i > 0. \tag{1}$$

Variables w_i give weights, how long time queues i are served per total time. Therefore, delay d_i in the queue i is actually

$$d_i = \frac{\Delta t_i}{w_i}. \tag{2}$$

Without loss of generality, only non-empty queues are considered, and therefore,

$$w_i \neq 0, \quad i = 1, \ldots, m, \tag{3}$$

where m is number of service classes. When one queue becomes empty, $m \to m - 1$.

Bandwidth or bit rate is formulated as follows. Let the processing time of the data be T [seconds/bit] in the packet scheduler. There are N_i connections or packets in the class i. Let us denote the packet size b_{ij} [bits] or [kbytes] in the class $i = 1, \ldots, m$ and the connection $j = 1, \ldots, N_i$. It is easy to see that bandwidth of the packet (i, j) is

- linearly proportional to the packet size b_{ij},
- linearly proportional to the weight w_i,
- inversely proportional to the processing time T, and
- inversely proportional to the total sum of the packet lengths $b_{ij}, j = 1, \ldots, N_i$ because other packets occupy the same band in a time-divided manner.

Therefore, the expression for the bandwidth is

$$B_{ij}[bits/s] = \frac{b_{ij} w_i}{N_i E(b_i) T} = \frac{b_{ij} w_i}{N_i E(b_i)} = \frac{b_{ij} w_i}{\sum_{l=1}^{N_l} b_{il}} \tag{4}$$

where the processing time T can be scaled $T = 1$, without loss of generality. Here

$$E(b_i) = \frac{1}{N_i}\sum_{j=1}^{N_i} b_{ij} \tag{5}$$

is mean packet length in the class *i*.

3. PRICING MODELS AND REVENUE MAXIMIZATION

We concentrate on the pricing and fair resource allocation from the point of view of the customers. On the other hand, from the point of view of the service provider, we try to maximize revenue. First, we introduce the concept of *pricing functions*. In the scope of our study, there are two QoS parameters, namely delay and bandwidth. Therefore, two separate pricing functions are defined.

A. General pricing function

Let the general pricing function be $f = f(w_1, \ldots, w_m)$. That means, f depends on the QoS parameters - in our study, delay and bandwidth - while QoS parameters depend on the weights w_i of the scheduler. Let the constraint for the weights be

$$\sum_{i=1}^{m} w_i = 1, \quad w_i > 0. \tag{6}$$

Revenue has the Lagrangian form

$$R = f + \lambda(1 - \sum_i w_i) \tag{7}$$

Derivative with respect to the weights is

$$\frac{\partial R}{\partial w_i} = \frac{\partial f}{\partial w_i} - \lambda = 0. \tag{8}$$

Then - because $\Sigma_i w_i = 1$:

$$\lambda = \lambda \sum_i w_i = \sum_i \lambda w_i = \sum_i \frac{\partial f}{\partial w_i} w_i. \tag{9}$$

Thus derivative of the revenue is

$$\frac{\partial R}{\partial w_i} = \frac{\partial f}{\partial w_i} - \sum_i \frac{\partial f}{\partial w_i} w_i. \tag{10}$$

Let us check the correctness of the derivative by direct substitution. Substitute Eq. (9) into Eq. (7). Then we obtain

$$R = f + \sum_k \frac{\partial f}{\partial w_k} w_k (1 - \sum_l w_l), \tag{11}$$

and remember the constraint (6). Then we obtain

$$\begin{aligned} \frac{\partial R}{\partial w_i} = \frac{\partial f}{\partial w_i} & \\ & + \sum_k \frac{\partial^2 f}{\partial w_i \partial w_k} w_k \\ & + \frac{\partial f}{\partial w_i} \\ & - \sum_k \frac{\partial^2 f}{\partial w_i \partial w_k} w_k \sum_l w_l \\ & - \frac{\partial f}{\partial w_i} \sum_l w_l \\ & - \sum_k \frac{\partial f}{\partial w_k} w_k \\ & = \frac{\partial f}{\partial w_i} - \sum_k \frac{\partial f}{\partial w_k} w_k \end{aligned} \tag{12}$$

But this is the same as (10) in the space $\{\Sigma_i w_i = 1\}$.

B. Pricing models for delay and bandwidth

The drawback in the previous scenario is that the algorithm developed from the gradient of the revenue is too complicated for fast implementation. Here we present the modified revenue criterion. Let the revenue be presented in the constrained form

$$R = f + \sum_k w_k (1 - \sum_l w_l) \tag{13}$$

Then

$$\frac{\partial R}{\partial w_i} = \frac{\partial f}{\partial w_i} + 1 - \sum_l w_l - \sum_k w_k = \frac{\partial f}{\partial w_i} - 1 = 0 \tag{14}$$

under the constraint $\Sigma_i w_i = 1$. It is seen from the criterion that the sum of the weights acts as the penalty term. If we consider linear pricing scenario, where revenue is decreasing as a function of the delay, and increasing as a function of the bandwidth, we get the pricing functions

$$f_1(d) = -\sum_i r_i d_i, \tag{15}$$

where r_i is the penalty factor for the delay in the class *i* in the pricing function, and *d* is the delay for the scheduler. On the other hand,

$$f_2(B) = \sum_i e_i B \tag{16}$$

is the pricing function for bandwidth *B*, where e_i is the pricing factor for the service class *i*. Total revenue can be expressed in the form

$$R = f_1 + f_2 + \sum_k w_k (1 - \sum_l w_l). \tag{17}$$

For all the connections, the pricing function can be presented in the form

$$f = -\sum_{i=1}^{m} \frac{N_i r_i \Delta t_i}{w_i} + \sum_{i=1}^{m} e_i w_i \tag{18}$$

$$\frac{\partial f}{\partial w_i} - 1 = \frac{N_i r_i \Delta t_i}{w_i^2} + e_i - 1 = 0 \tag{19}$$

Then the closed form approximation for the weights is as follows:

$$w_i = \frac{\sqrt{\dfrac{N_i r_i \Delta t_i}{1 - e_i}}}{\sum_k \sqrt{\dfrac{N_k r_k \Delta t_k}{1 - e_k}}} \tag{20}$$

4. EXPERIMENTS

In the first experiment, we use static traffic profile to justify the performance of the closed form algorithm. The parameters are as follows:

- N = 15; 25; 70 number of connections in the gold, silver, and bronze classes, respectively.
- Pricing factors for delays are r = 30; 15; 5 for gold, silver, and bronze classes, respectively.
- Time delays of the buffers are _t = 10; 25; 50 for gold, silver, and bronze classes, respectively.
- Vertical price shifting parameters are k = 10000; 5000; 3000 for gold, silver, and bronze classes, respectively.
- Pricing factors for bandwidths are e = 30; 25; 22.

The results are as follows. Revenue for closed form algorithm = 396950, and revenue for brute-force method = 397290. Then, the revenue closed form/brute-force = 99.91 %. Thus, we can conclude that the revenue obtained by the closed form approximation is very near compared to the revenue obtained by the optimal brute-force method.

In the second experiment, the parameters are as follows:

- N = 30; 50; 100 number of connections in the gold, silver, and bronze classes, respectively.
- Pricing factors for delays are r = 40; 20; 10 for gold, silver, and bronze classes, respectively.
- Time delays of the buffers are _t = 10; 25; 50 for gold, silver, and bronze classes, respectively.
- Vertical price shifting parameters are k = 10000; 5000; 3000 for gold, silver, and bronze classes, respectively.
- Pricing factors for bandwidths are e = 30; 25; 22.

Revenue for closed form algorithm = 607730, and revenue for brute-force method = 608680. Then, the revenue closed form/brute-force = 99.84 %.

In the third experiment, the parameters are as follows:

- N = 100; 50; 15 number of connections in the gold, silver, and bronze classes, respectively.
- Pricing factors for delays are r = 50; 25; 15 for gold, silver, and bronze classes, respectively.
- Time delays of the buffers are _t = 20; 50; 25 for gold, silver, and bronze classes, respectively.
- Vertical price shifting parameters are k = 10000; 5000; 3000 for gold, silver, and bronze classes, respectively.
- Pricing factors for bandwidths are e = 30; 25; 22.

Revenue for closed form algorithm = 882440, and revenue for brute-force method = 883850. Then, the revenue closed form/brute-force = 99.84 %.

5. DISCUSSION AND CONCLUSIONS

Here we discuss the results and conclude the work:

- We considered delay guarantee and bandwidth allocation of communications network.
- Pricing scheme was linear for both QoS parameters.
- We developed novel constrained optimization approach which resembles Lagrangian approach.
- The approach yielded fast closed form approximation for optimizing the revenue of the service provider.
- On the other hand, the algorithm gives fair resource allocation.

General conclusion is the our approach makes possible for all people - including the people of modest means - to use communication services by using different pricing classes.

REFERENCES

[1] 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 Transactions on Automatic Control, Vol. 47 Issue: 6*, June 2002 pp. 917–930.

[2] E. Altman, T. Basar, T. Jiminez and N. Shimkin, "Competitive routing in networks

with polynomial costs". *IEEE Transactions on Automatic Control, Vol. 47 Issue: 1*, Jan. 2002, pp. 92–96.

[3] T. Basar, R. Srikant, "Revenue-maximizing pricing and capacity expansion in a many-users regime". *Proc. of IEEE INFOCOM 2002*, Vol. 1, 2002, pp. 294–301.

[4] J. C. R. Bennett and H. Zhang, "WF2Q: Worst-case Fair Weighted Fair Queueing". *Proc. of INFOCOM 1996*, 1996, pp.

[5] H. M. Chaskar and U. Madhow, "Fair Scheduling with Tunable Latency: A Round Robin Approach". *Proc. of IEEE Globecom 1999*, 1999, pp.1328–1333.

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

[7] A. Demers, S. Keshav, and S. Shenker, "Analysis and Simulation of a fair queueing algorithm". *Proc. of IEEE SIGCOM 1989*, 1989, pp.1–12.

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

[9] T. Hämäläinen and J. Joutsensalo, "Link Allocation and Revenue Optimization for Future Networks", *Proc. of IEEE Globecom 2002*, Nov. 2002, Taiwan.

[10] Jong-Seon Kim; Lee, D.C., "Weighted round robin packet scheduler using relative service share". *Proc. of IEEE MILCOM 2001, Vol.2*, 2001, pp. 988–992.

[11] J. Joutsensalo and T. Hämäläinen, "QoS-Aware Adaptive Pricing for Network Services". *Proc. of IEEE Globecom 2001, Vol. 4*, Nov. 2001, pp. 2455–2459.

[12] J. Joutsensalo and T. Hämäläinen, "Optimal Link Allocation and Revenue Maximization". *Journal of Communications and Networks, Vol. 4, No. 2*, June 2002, pp 136–147.

[13] J. Joutsensalo, T. Hämäläinen, Mikko Pääkkönen, and Alexander Sayenko, "Adaptive Weighted Fair Scheduling Method for Channel Allocation". to be published in *Proc. of IEEE ICC 2003*, Anchorage, Alaska, 2003.

[14] F. P. Kelly, "Charging and rate control for elastic traffic". *European Transaction on Telecommunication., vol. 8*, 1997, pp. 33—37.

[15] 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*, 1998, pp. 237-–252.

[16] F. P. Kelly, "Notes on effective bandwidths". *in Stochastic Networks: Theory and Applications*, S. Zachary, I. B. Ziedins, and F. P. Kelly, Eds. London, U.K.: Oxford Univ. Press, 1996, vol. 9, pp. 141-–168.

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

[18] Y. Korilis, A. Lazar, and A. Orda, "Architecting noncooperative networks". *IEEE Journal on Selected Areas in Communications, vol. 13*, 1995, pp. 1241—1251.

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

[20] S. Kunniyur and R. Srikant, "End-to-end congestion control: Utility functions, random losses and ECN marks", *Proc. of IEEE INFOCOM 2000*, 2000, Tel Aviv, Israel, pp. 1323-–1332.

[21] R. J. La and V. Anantharam, "Utility-Based Rate Control in the Internet for Elastic Traffic". *IEEE/ACM Transactions on Networking, Volume: 10 Issue: 2*, April 2002, pp. 272–286.

[22] R. J. La and V. Anantharam, "Optimal routing control: a game theoretic approach". *Proc. of the 36th IEEE Conference on Decision and Control*, San Diego, CA, December 1997.

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

[24] J. K. MacKie-Mason and H. R. Varian, "Pricing the Internet, in Public Access to the Internet". *B. Kahin and J. Keller, Eds. Englewood Cliffs, NJ: Prentice-Hall*, 1994.

[25] N. Matsufuru and R. Aibara, "Efficient Fair Queueing for ATM Network using Uniform Round Robin". *Proc. of INFOCOM 1999*, 1999, pp. 389– 397.

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

[27] 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.

[28] I. Ch. Paschalidis and Yong Liu, "Pricing in multiservice loss networks: static pricing, asymptotic optimality and demand substitution effects". *IEEE/ACM Transactions on Networking, Volume: 10 Issue: 3*, June 2002 pp. 425–438.

[29] 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 Transactions on Networking, vol. 1, no. 3*, June 1993, pp. 344–357.

[30] A. Orda, R. Rom, and N. Shimkin, "Competitive routing in multiuser communication networks". *IEEE/ACM Transactions on Networking, vol. 1(5)*, Oct. 1993, pp. 510—521.

[31] J. Joutsensalo, T. Hämäläinen, M. Paakkonen, and A. Sayenko: "QoSand Revenue Aware Adaptive Scheduling Algorithm", in *Journal of Communications and Networks*, Vol. 6, No.1 March 2004, pp. 68-77.

[32] Jyrki Joutsensalo, Timo Hämäläinen, Kari Luostarinen, and Jarmo Siltanen, “Adaptive Scheduling Method for Maximizing Revenue in Flat Pricing Scenario.” *AEU - International Journal of Electronics and Communications*, vol. 60, issue 2, February 2006, pp. 159-167.

BIOGRAPHY

Jyrki 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, 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.

Timo Hämäläinen received the B.Sc in automation engineering from the Jyväskylä Institute of Technology in Finland on 1991 and the M.Sc and Ph.D degrees in telecommunication from Tampere University of technology and University of Jyväskylä, Finland in 1996 and 2002, respectively. Currently he is Professor of Telecommunications at the University of Jyväskylä and his research interests include traffic engineering and QoS in wired and wireless networks.

Jarmo Siltanen received the B.Sc in computer and communication engineering from the Jyvskylä Institute of Technology in Finland on 1996. The M.Sc and Licentiate in Philosophy degrees in telecommunication he received from University of Jyvskylä Finland in 1998 and 2003, respectively. His Ph.Lic works studied Dynamic Resource Allocation and Quality of Service in networks. His current research interests include traffic engineering and QoS in wired and wireless networks.

M.Sc Isto Kannisto is marketing head of TeliaSonera and his research interest include ICT-marketing, traffic engineering and QoS in wired and wireless networks.