https://doi.org/10.26552/com.C.2001.4.5-8

Péter Fodor – Ádám Marquetant *

POSÚDENIE STRATÉGIÍ TARIFIKÁCIE PRUŽNÉHO PREVÁDZKOVÉHO ZAŤAŽENIA NA KOMUNIKAČNOM SPOJI

EVALUATION OF PRICING STRATEGIES OF ELASTIC TRAFFIC ON A SINGLE COMMUNICATION LINK

Budúce siete s integrovanými službami, podporujúce viaceré triedy prevádzkového zaťaženia, budú vyžadovať zákaznícke stratégie riadenia prístupu a zdieľania frekvenčného pásma, ktoré splnia rôzne požiadavky na zaručenú kvalitu služieb (QoS) alebo dátový tok a pružnosť služieb. Výsledky súčasného výskumu ukazujú, že je dôležité overiť riadenie prístupu volaní (CAC) pre pružné prevádzkové zaťaženie, nakoľko algoritmus CAC umožňuje napríklad zabrániť reláciám TCP pre nadmernú degradáciu priepustnosti [1], [2]. Aby bolo možné pre pružné volania určiť optimálny algoritmus CAC, overili sme na základe modelu spoja zavedeného v [2] pomocou Markovovej teórie rozhodovania rôzne stratégie tarifikovania. Ukážeme, že optimalizácia CAC maximalizuje nielen príjmy, ale tiež zvyšuje pravdepodobnosť blokovania prevádzky tokov s vysokou prioritou a QoS pružného prevádzkového zaťaženia, ak je tarifikačná funkcia použiteľná.

The future integrated service networks supporting multiple traffic classes will require customized admission control and bandwidth sharing strategies, which meet the diverse needs of QoS (Quality of Service)-assured (stream) and best-effort (elastic) services. Recent research results indicate that it is meaningful to exercise call admission control (CAC) even for elastic (best-effort) traffic, because CAC algorithms provide a means to prevent e.g. TCP sessions from excessive throughput degradations [1], [2]. Based on a model of a single link introduced in [2], we evaluate different pricing strategies assigned to elastic calls by determining an optimal CAC using Markov Decision theory. We will show that optimizing CAC not only maximizes average revenue, but also improves blocking probability of high priority stream traffic and QoS of elastic traffic as long as appropriate pricing functions are applicable.

1. System Model

We formulate our system model following the approach described in [1] and [2].

Traffic Model We investigate a single link, to which two types of traffic classes offer load: *stream* and *elastic*. Stream traffic is supposed to represent a service with strict QoS guarantees (e.g. VoIP in IP or CBR in ATM), while elastic traffic models best effort-like services (e.g. TCP in IP, ABR in ATM). Stream calls are described by their arrival rate λ_1, departure rate μ_1 and peak rate B_1, and elastic calls by their call arrival rate λ_2, their ideal departure rate μ_2, their peak rate B_2 and minimum rate $r_{min}*B_2$ ($0 \le r_{min} \le \le 1$). Both types of calls arrive according to independent Poisson processes and the holding time for stream flows is exponentially distributed with mean μ_1^{-1}. In case of elastic calls the number of bits to transfer is exponentially distributed with mean $B_2*\mu_2^{-1}$. By ideal departure rate we mean that the actual *service ratio* $r(t)$ of elastic calls in progress may fluctuate between r_{min} and 1, thus the service time increases accordingly. All elastic connections in progress on the link share the available bandwidth equally among them [4]. A newly arriving call will be accepted if there is enough free capacity on the link by the compression of elastic flows (elastic flows can be compressed down to $r_{min}*B_2$). If the available free capacity on the link is smaller than the minimum rate of the new call, then the flow will be rejected. When a flow departures elastic calls inflate their bandwidth consumption up to B_2.

System Description Let C denote the link capacity. The system under investigation (with the above assumptions regarding the arrival processes and holding times) is a Continuous Time Markov Chain (CTMC) whose state is uniquely characterized by $(n_1(t), n_2(t))$ where $n_1(t)$ is the number of stream calls and $n_2(t)$ is the number of elastic calls on the link at time t ($0 \le n_1(t) \le \lfloor C / B_1 \rfloor$, $0 \le n_2(t) \le \lfloor (C - n1(t)) / (r_{min} * B_2) \rfloor$). Let the state space S. The vector $(n_1(t), n_2(t))$ uniquely specifies what service ratio $r(t)$ of in-service elastic calls receive $r(t) = \min[1; (C - n_1(t) * B_1) / / (n_2(t) * B_2)]$. In order to obtain the performance measure of this system we need to determine the CTMC's generator matrix Q and its steady state solution, $\underline{P}$. The non-zero transition rates of generator matrix are:

* **Péter Fodor, Ádám Marquetant**
High Speed Networks Laboratory, Department of Telecommunications and Telematics, Budapest University of Technology and Economics, H-1117, Pázmány Péter sétány 1/D, Budapest, Hungary, Tel.: +36-1-463 2187 Fax: +36-1-463 3107
e-mail: fodorp@ttt-atm.ttt.bme.hu, marquet@ttt-atm.ttt.bme.hu

$$q(n_1, n_2; n_1 + 1, n_2) = \lambda_1$$

$$q(n_1, n_2; n_1 - 1, n_2) = n_1 \cdot \mu_1$$

$$q(n_1, n_2; n_1, n_2 + 1) = \lambda_2$$

$$q(n_1, n_2; n_1, n_2 - 1) = n_2 \cdot r_2(n_1, n_2) \cdot \mu_2$$

according to call arrival or departure. The state probability vector has to satisfy $\sum_{j \in S} p_i = 1$, and $\underline{PQ} = \underline{0}$. Let $\underline{B_i}$ denote the blocking probability vector of traffic class *i*. $\underline{B_i}$ contains those states where a newly arriving call from traffic class *i* is blocked.

$$\underline{B}_1 = \{(n_1, n_2) \in S : (n_1 + 1, n^2) \notin S\} \text{ - stream}$$

$$\underline{B}_2 = \{(n_1, n_2) \in S : (n_1, n_2 + 1) \notin S\} \text{ - elastic}$$

Then the blocking probability of traffic class i is

$$P_i = \sum_{(n_1, n_2) \in B_i} p_{(n_1, n_2)}. \quad (1)$$

To get the average holding time of elastic calls we need to know the mean number of elastic flows on the link from

$$E[n_2] = \sum_{(n_1, n_2) \in S} n_2 \, p_{(n_1, n_2)} \quad (2)$$

where n_2 is the number of elastic calls in all state. From Little's formula the mean time an elastic call spends in the system is $E[T_2] = E[n_2] / (\lambda_2 \cdot (1 - P_2))$ and the average service ratio of elastic flows is $r_{avg} = 1 / (E[T_2] \cdot \mu_2)$.

Pricing Model To represent that stream and elastic calls are of different value to the provider we also assume that both types of calls generate revenue that is a function of the occupied bandwidth. The link-wide instantaneous revenue accumulation rate $\rho(t)$ is given by

$$\rho(t) = n_1(t) * B_1 + n_2(t) * B_2 * \varphi(B_2, r(t), r_{min}) \quad (3)$$

We assume that a unit stream bandwidth generates revenue with a unit rate, while a unit elastic bandwidth generates revenue with a rate proportional to φ.

Optimal CAC Policy based on Markov Decision Theory The simplest call admission policy (CAC) may be the one, which admits a new call whenever the link is capable to accommodate it (i.e. by compressing all elastic flows down to their minimum ratio r_{min}). Note, that we will refer to this kind of CAC 'no Markov Decision' (*no_MD*) in Section 4. We argue that there is a need to apply more sophisticated CAC policies with the following two reasons. First, the provider is seeking after to increase its income, therefore it is straightforward to price stream flows requiring strict QoS guarantees higher and prefer them whenever both stream and elastic flows aspire for admission. Second, users generating stream flows expect better service deservedly for their money.

We aim at finding a CAC policy that assigns to each system state a decision whether to admit or reject arriving stream and elastic calls so that the long-term revenue is maximized. To achieve our goals we apply Markov Decision theory [3], which algorithmically takes into account the revenue generation rate of different system states and yields the optimal solution in a finite number of steps.

2. Pricing Strategies

Formula (3) allows a multitude of pricing strategies to apply to elastic calls. We present two of them, which we think are relevant in the context of optimizing CAC to achieve maximum possible revenue for the provider and improve QoS and blocking probability of stream flows. First of all we introduce our underlying assumptions.

First, the price of a stream call's unit bandwidth should be higher than that of an elastic call to be able to satisfy strict QoS requirements by ensuring that only a fraction of users will claim to those services. Secondly, it is beneficial to price elastic calls requiring larger minimal service ratios r_{min} higher. (These calls have a larger percentage of their bandwidth guaranteed and are more like to stream calls.) Otherwise, subscriber may spare money by offering a wider band elastic call instead of an expensive stream call ($B1 \approx r_{min} * B_2$). We have found two simple pricing strategies fulfilling the above requirements (see Fig. 1).

2.1. 'Linear' Pricing

The revenue generation factor of elastic traffic is directly proportional to the service ratio. (Represented as a line on Fig. 1, which would begin from the origin if we could decrease the service ratio down to zero) ($\varphi_{lin}(t) = rev_{lin} * r(t) = k * r_{min} * r(t)$). This strategy fulfills both criteria mentioned above, provided constant k falls into range ($0 < k \leq 1/r_{min}$). Note that different r_{min} values entail lines with different gradient on Fig. 1.

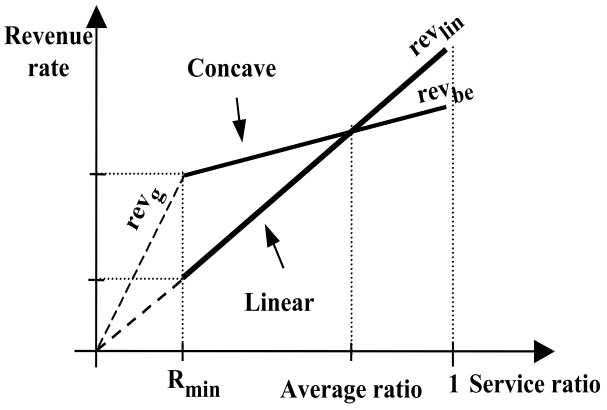

Fig. 1

2.2. 'Concave' Pricing

At the second solution this line is shifted up and its gradient is smaller than that of the other. In this case the revenue rate has two parts. One of them is a guaranteed rate, what the customer always has to pay. The best effort part has linear increment, which has to be paid when the service ratio is greater than the minimum service ratio.

($\varphi_{con}(t) = rev_g * r_{min} + rev_{be} * (r(t) - r_{min})$), which meets the above mentioned requirements if $0 \leq rev_{be} \leq rev_g \leq 1$. So the curve must be concave, hence the name.

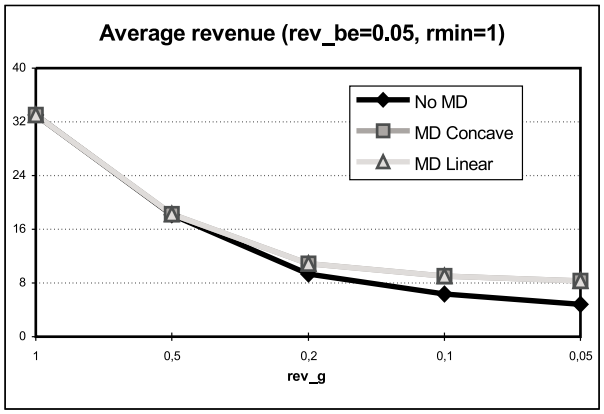

Fig. 2

3. Numerical Results

To evaluate the pricing strategies we investigated a single link of capacity $C = 40$ Mbps. Stream calls with bandwidth demand $B_1 = 8$ Mbps arrive with intensity $\lambda_1 = 1\ s^{-1}$, while elastic calls with bandwidth demand $B_2 = 1$ Mbps (peak) arrive with intensity $\lambda_2 = 8\ s^{-1}$, where r_{min} decreases from 1.0 down to 0.2. The holding times of two traffic classes are assumed to be exponential with mean values $1/\mu_1 = 1$ s and (ideally) $1/\mu_2 = 4$ s. Aside from r_{min}, our moving parameters are rev_g (assuming values 0.05, 0.1, 0.2, 0.5, 1) and rev_{be} (assuming values 0.05, 0.2, 0.4). The offered traffic to the link is equal to its capacity in all our measurements, hereby we were modeling a good provisioned network (link). Our main measures of interest are long-term average revenue, stream and elastic class blocking probability and average elastic service ratio (r_{avg}) and holding time. Fig. 2, Fig. 3, Fig. 4, Fig. 5 and Fig. 6 show these measures as the function of guaranteed revenue factor (rev_g) for a fixed best-effort revenue factor (rev_{be}) and minimal service ratio (r_{min}) of elastic calls.

On Fig. 2, Fig. 3, Fig. 4, Fig. 5 and Fig. 6 we compare three kinds of CAC algorithms labeled *'no MD', 'MD concave'* and *'MD linear'*. *'no MD'* stands for the simple CAC that does not take into account revenue generation rates of calls, but works on an is-enough-bandwidth basis (see Section 2.). The other two CAC's are outcome applying the Markov decision optimization in the same environment (with the same elastic r_{min}) as in case of the respective 'no MD' CAC, using respective pricing strategies as input to the optimization. To be able to compare the *linear* and *concave* pricing strategy we first apply the 'no MD' CAC and based on the resulting average elastic service ratio (r_{avg}) we calculate the gradient of the linear strategy (rev_{lin}) as the function of concave parameters as follows: $rev_{lin} = (rev_g * r_{min} + (r_{avg} - r_{min}) * rev_{be}) / r_{avg}$ (see Fig. 1). Thus, under 'no MD' both strategies will produce the same average revenue (they 'offer' the same amount of revenue to the network).

Fig. 2 shows the average revenue of the link in function of guaranteed revenue factor (rev_g) of elastic calls. Naturally, the greater is the rev_g, the greater is the long-term average revenue. This figure also reveals that Markov Decision is only capable of effective increase average revenue when elastic flows are much lower priced (rev_g is close to 0.05), however, at the price of much higher elastic blocking probabilities. Comparing two pricing policies in terms of guaranteed revenue rate of elastic calls there is no sufficient variance.

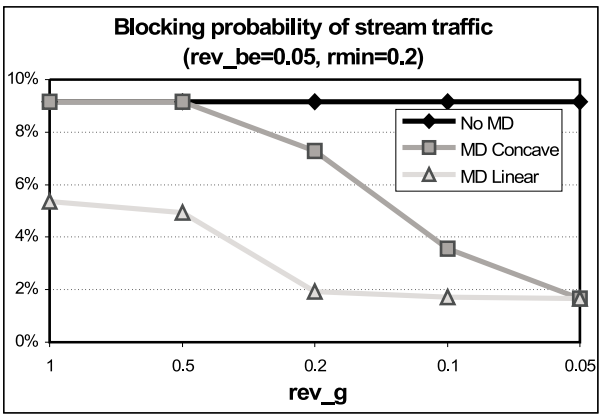

Fig. 3

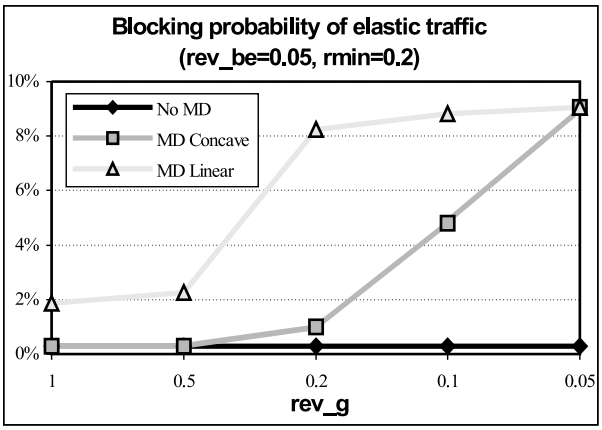

Fig. 4

Fig. 3 and Fig. 4 show interesting results. Even if there is no considerable average revenue increase with MD, the blocking probabilities are much more effected. Without MD, stream calls have

much higher blocking probability than elastic calls. MD, however, decreases the blocking probability of stream traffic, while that of elastic flows will be increased. We just mention that by applying MD elastic average service rate (r_{avg}) has been increased providing better QoS to elastic flows (Fig. 5).

In linear pricing case the blocking probability of stream calls decreased more than when using the concave pricing policy. Analogously, blocking probability of elastic calls increased more by the linear policy. These results are caused by two phenomena. The MD algorithm is blocking elastic flows even if there was place for them on the link to admit newly arriving stream calls with higher probability. Elastic calls can expand more in the meantime therefore the average rate of service ratio shifts right. In linear pricing case this expansion generates the same per unit bandwidth revenue as if we admitted more elastic calls with smaller service ratio. However, in case of concave pricing, increase of average service ratio would result in lower per unit-bandwidth revenue, since the gradient of its revenue generation factor is smaller. So, the smaller is the service ratio of elastic flows priced concave, the greater is the long-term average revenue of the link. Therefore the algorithm considers the probability of arriving a new great-revenue stream call of less value, which causes greater blocking probability for stream traffic.

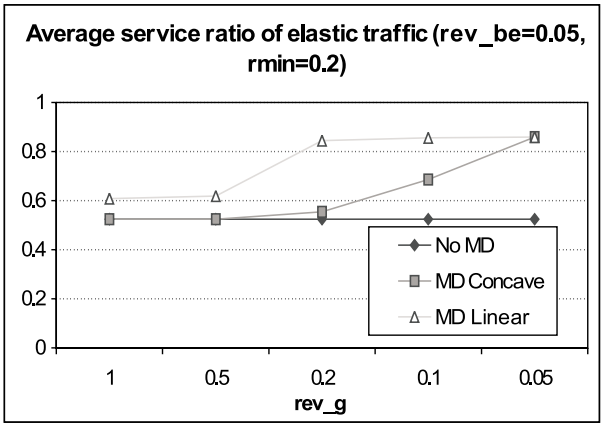

Fig. 5

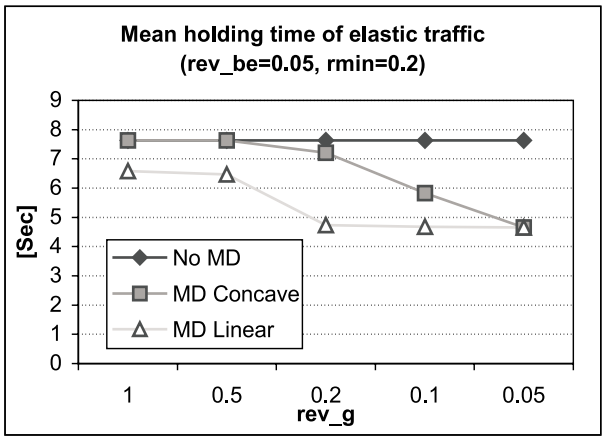

Fig. 6

Fig. 5 shows the average service ratio of elastic flows. In accordance with the above mentioned phenomena if the revenue factor of guaranteed part is much higher than that of the best effort part, in case of concave pricing strategy MD tries to keep the average service ration of elastic calls near minimum service rate. In the linear pricing case this ratio increased much more since more elastic calls are blocked. If the guaranteed part is lowly priced then the effect of second phenomenon diminishes, therefore the concave curve approximates to linear. Fig. 6 shows the mean holding time of elastic calls, which is really for subscribers. It is inversely proportional to the average service rate, therefore in linear pricing case the mean time that an elastic call spends in the system is smaller than in concave pricing case.

4. Conclusion

In this paper we studied pricing policies for elastic flows to be applied on a communication link accommodating stream and elastic calls. First, we showed a model that is able to describe the dynamics of the link and then presented elastic pricing policies which in combination with Markov Decision optimization can yield better blocking probability measures for high priority stream traffic and maximize revenue of network provider.

5. References

[1] Rudesindo Núñez Quejia, Hans van den Berg, Michel Mandjes: *Performance Evaluation of Strategies for Integration of Elastic and Stream Traffic*

[2] G. Fodor, E. Nordstrom and S. Blaabjerg: *Revenue Optimization and Fairness Control of Priced Guaranteed and Best Effort Services on an ATM Transmission Link*, ICC'98, vol 3, pp. 1696-1705, Atlanta, GA, USA, 1998

[3] Henk C. Tijms: *Stochastic Models - An Algorithmic Approach*, John Wiley & Sons, Chichester, 1984

[4] Shigang Chen, Klara Nahrstedt: *An Approach to Pricing and Resource Sharing for Available Bit Rate (ABR) Services*, in Proceedings of Euro-Parallel and Distributed Systems Conference (Euro-PDS '98), Vienna, Austria, July 1998, pp.163-168