Charging Schemes for Network Systems

Marian Dobre, Marius Dorian Zaharia, Dorin Irimescu

University POLITEHNICA Bucharest, Computers Department,
str. Splaiul Independentei, No. 313, 77206 – Bucharest, Romania
E-mail: {mdobre,zaharia,irimescu}@cs.pub.ro

Alexandru Soceanu

Fachhochschule Regensburg, Fachbereich Informatik/Mathematik,
Prüfeninger Sraße 58, 93049 Regensburg, Germany
E-mail: alexandru.soceanu@informatik.fh-regensburg.de

Abstract: The paper describes some models for charging network services, emphasizing the case of usage based charging. This is the first necessary step in designing and implementing of a billing system. The next step will be to introduce into the network data acquisition points, based on cheap machines (PC-s).

Keywords: network, network service, game theory, bandwidth, charging.

1. Introduction

The extraordinary increase of the number of new INTRANET networks as well as the integration of the already existent local area networks in INTERNET determined the growth of the network maintenance costs as well as of the various network resources (buffers, routers, switches). This situation requires new methods that could allow a fair management of the network utilization.

The recent technological developments – i.e. ATM (Asynchronous Transfer Mode) networks allow to a user to specify the necessary bandwidth for a certain application and a minimal acceptable level for the quality of the required network services. Various schemes that establish the load of different network communication links, or other network resources, and the associated prices could be imagined.

A user oriented billing system:

- should not require the modification of the existing applications or protocols
- should accomplish the real time data acquisition and online reporting of network usage in order to determine the part of the price due to the network congestion
- in case of applications that necessitate cooperation between users or sharing the information and resources between network sites, a charging system should identify the cooperating users, in order to charge correctly the services offered to each of them. The system must be able to detail user identities and explain charges.

As generally accepted [1], [2], [4], [10], four types of charges can be considered:

- The access charge, which is independent of the number of connections of a user i.e. independent of the network usage. This type of fixed cost covers the hardware or software maintenance and administration.
- The usage based charging depending on the number of connections made by the user. It may be calculated as a function of the amount of the transferred data, the duration and the length (end to end distance) of a connection.
- Congestion based pricing is function of the load ratio of the network at the moment when the user initiates the connection. This type of

- charging depends on the network status (the charge grows when the network is congested). In a statistical multiplexing network, the congestion could contribute to increase the queuing delays and the loss of packets due to an overflow of the buffer capacities. In order to prevent the network congestion, a special congestion fee could be used to eliminate the “less valuable connections”.
- The charge of the quality of service (QoS) that depends on the quality of resources (QoR) requested by the user. This is specific to new, evoluated network devices, which could provide different service qualities and to new applications that require large bandwidths.

The organization of the paper is as follows. In chapter 2, the ways the usage of a network resource might be defined and measured are briefly discussed. In particular, the concept of effective bandwidth is introduced as a mean of such a measure. In chapter 3 a mathematical formulation is introduced for two different pricing methods. Chapter 4 presents some aspects of selecting the price so that the global aim of system optimality and the users’ individual aims are harmonized.

2. The effective bandwidth model

In the case of broadband networks, the duration of a connection does not have the same importance as in phone networks. The resources are dynamically allocated and the main factors influencing the amounts of resources to be allocated are the different rates at which the traffic is generated and transmitted. These rates determine the volume of the traffic (the product of a connection duration and rate).

The concept of effective bandwidth has been developed by several authors [3], [4], [7], [9], as a measure for the usage of a resource shared by sources with various traffic characteristics. The effective bandwidths could be used to find the usage-based component of a charge.

Let NR be the number of different resource types, n_j the number of the traffic sources of type j and X_{ji} the fraction of the total resource capacity that is loaded by the j-th source of type i ($1 \le i \le n_j$, $1 \le j \le NR$). X_{ji} are considered random variables and have identical distributions. The effective bandwidth at the resource due to a source of type j is given by:

$$a_j(s) = \frac{\log(E[\exp(s X_{ji})])}{s} \qquad (1)$$

$a_j(s)$ is the common logarithmic moment generating function of the random variable X_{ji} ($1 \le i \le n_j$). If C is the resource capacity and the total load due to the above mentioned traffic sources is:

$$S = \sum_j \sum_i X_{ji} \qquad (2)$$

then the probability that S exceeds C is bounded by $e^{-\gamma}$ (where $\gamma \ge 0$):

$$P(S \ge C) \le e^{-\gamma} \qquad (3)$$

if there is some $s \ge 0$ such that:

$$\sum_j n_j a_j(s) \le C - \frac{\gamma}{s} \qquad (4)$$

In case of an unbuffered resource that has the peak rate of traffic generation h and the mean rate m:

$$a(s) = \frac{\log\left[1 + \frac{m}{h}(e^{sh} - 1)\right]}{s} \qquad (5)$$

and

$$\lim_{s \to 0} a(s) = m, \qquad \lim_{s \to \infty} a(s) = h \qquad (6)$$

corresponding to situations when C is very large ($s \to 0$) or h is large with respect to C ($s \to \infty$).

In order to determine the usage based charge, the effective bandwidth of a connection could be a-posteriori estimated, based on measurements made for this connection. This could cause errors because the network has to reserve resources at connection initiation time.

Other method of estimating the effective bandwidth uses the history of the data measured on connections of the same type. A category of a-priori information could be a set of mutually agreed parameters specified in the traffic contract, for example, in case of ATM networks the peak rate and the sustainable cell rate could replace *h* and *m* in order to determine the effective bandwidth of a source.

3. Control of the network usage through pricing methods

Some results in the game theory can be used to model the charging schemes. If the whole network is considered as a "global" resource which is shared by n users and S is the set of services which could be requested by each user, then the net outcome of user i (1 ≤ i ≤ n) is:

$$u_i(\sigma) = v_i(\sigma) - c_i(\sigma) \qquad (7)$$

where $c_i(\sigma)$ is the amount by which user *i* is charged, and $v_i(\sigma)$ is the total gain acquired by the user *i*.

$\sigma=\{\sigma_j \mid 1 \le j \le n\}$ is a collection of requests sent by the users to the network at a given time.

A measure of the degree of efficient usage of the network could be the total gain of all users i.e. a vector $\sigma^{\max}$ of requests that maximizes the sum of v_i-s:

$$\sum_i v_i(\sigma^{\max}) \ge \sum_i v_i(\sigma) \quad for \quad \forall \sigma \in S^n \qquad (8)$$

This measure, named "social welfare function" is suitable to be used for modeling charging in public, nonprofit, research networks.

In case of private, profit-seeking firms, other pricing schemes should be developed. Generally each user reacts selfishly comparing the benefit of the service he is receiving with the price he must pay for the service. This kind of behavior could be modeled using game theory where:

- each "player" has limited knowledge about the whole system,
- each "player" has a choice between some possible actions,
- each "player" has an associated (individual) profit function
- the outcome for each "player" depends not only on his own actions but also on the actions of the others players too.

In order to solve this kind of non-cooperative game, each user *i* tries to maximize his individual net outcome u_i. The resulting vector of requests has the property that no user can increase his outcome by unilaterally changing his own request σ_i i.e.

$$u_i(\sigma^N) \ge u_i(\sigma^N \big|^i \rho_i), \forall i = 1,\dots,n, \forall \rho_i \in S \qquad (9)$$

where $\sigma^N|^i\rho_i$ is a vector obtained by replacing the i-th component of σ^N by ρ_i. This condition is characteristic to "Nash equilibrium point" for a non-cooperative game. The network service provider has a pricing policy, which could be described by a function

$$c: S^n \to R^n,\ \sigma \to (c_i(\sigma)) \qquad (10)$$

that assigns to each user *i*, having a vector σ of requests, a cost $c_i(\sigma)$, 1 ≤ i ≤ n.

The compatibility between the two above mentioned pricing methods: the global "social welfare" maximization (determination of $\sigma^{\max}$) and the game theory based pricing policy (determination of revenues associated to the network service provider and of the costs assigned to each customer) could be accomplished by choosing a pricing scheme so that: $\sigma^{\max}=\sigma^N$. For a network with free services:

$$(c_i(\sigma)=0,\ \forall \sigma \in S^n,\ \forall 1 \le i \le n) \qquad (11)$$

the condition for Nash equilibrium (9) requires that, for every user, the gain $v_i(\sigma)$ must be maximized i.e. the user should require the highest possible quality of service that is sufficient for his purpose.

If the strategies σ_i are interpreted as information transfer ratios, the Nash equilibrium would be possible only if none of the traffic participants would logically wish to increase his transfer ratio. In case of a limited network resource, an arbitrary increase of the data transfer rate associated to a user would reduce the network usage by other users that “build” together the Nash equilibrium. So, in this lastly mentioned case the social welfare approach is certainly non-optimal.

4. Price Selection

Unfortunately, the system optimality and individual aims of the users can be contradictory. The global aim of system optimality and the users’ individual aims shall be harmonized. Users who access a shared resource will only take into account their own benefits and costs from the resource usage, but not the costs of congestion (such as delay, packet loss or exclusion) resulting from this usage for the other customers. If these additional costs, however, would be included in the amount charged to the users, one could expect the global and individual aims to be harmonized. This is the main idea behind *congestion pricing*. In [4], [7], [8] one model is developed and assessed for this kind of charging.

The practical aspects of that problem concern the way this “game” can be controlled. We consider a network resource, which is shared by n users. We denote by:

x_i	the load placed on the resource by user *i*
$X := \sum x_i$	the total load placed on the resource by all the users i:=1,… ,n
K	the capacity of the resource
Y:=X/K	the utilization of the resource
$v_i(x_i,Y)$	the *gross utility* of user i. By this function the influence of the other users’ loads x_j, j≠i and the user's *i* load are summarized through the utilization Y.

An increasing utilization results in congestion and loss the users' utility. These dependencies can be represented by barrier function f, which determines the total amount of congestion costs caused by the congestion phenomena at a given degree of resource utilization. We assume the resource to be buffered. Then the resource utilization could be measured by the corresponding buffer occupancy:

$$f(b) = \frac{1}{B-b} - \frac{1}{B} \tag{12}$$

where: B is the number of buffer places

b is the number of occupied buffer places.

The condition to maximize the total gross utility [4] can be represented as:

$$\forall i = 1,\ldots,n; \qquad \frac{\partial v_i(x_i,Y)}{\partial x_i} - p = 0 \tag{13}$$

where *p* denotes the price.

The users’ utility functions will change in time. They will be equal to zero as long as the user is not willing to send at all. The time will be divided into pricing intervals of a certain length T. During each interval the users’ utility functions are assumed to be fixed. The mechanism to control the charging is further described.

a) The provider chooses an initial value for price p.
b) At the start of each pricing interval the provider announces to the users the price p for the following interval
c) The users choose their bandwidth for this interval according to condition (13)
d) At the end of the interval the optimal price p is calculated by the provider with respect to the values of the buffer occupancy and traffic load (in the ended interval)
e) The provider modifies the price correspondingly. Go to b).

Conclusions

The paper tried to describe how the usage of a network resource could be defined, measured, and controlled appropriately. The prices have to be chosen and modified so that system optimality to be harmonized with the users’ individual aims. The most important publications on usage based charging do not mention means to make predictions on the convergence behavior of different such algorithms. We will try to simulate the behavior of the algorithm presented here.

One problem is how to select the initial value of the price.

References

[1] R. Cocchi, S. Shenker, D. Estrin, L. Zang, “Pricing in Computer Networks: Motivation, Formulation, and Example”, *IEEE/ACM Transactions on Networking*, No. 1, 1993.

[2] R. Eddelm, N. Mc. Keown, P. P. Variya, “Billing users and pricing for TCP”, *IEEE Journal Selected Areas in Communications*, 13(7):1162-1175, 1995.

[3] A. I. Elwalid, D. Mitra, “Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks”, *IEEE Transactions on Networking*, No. 1, 1996.

[4] B. Hofmann, “Usage-based charging schemes for broadband networks”, Bericht Nr. 9703, Universität der Bundeswehr, München, Fakultät für Informatik, 1997.

[5] B. Hofmann, “Preisfestsetzung in ATM-Netzen”, Universität der Bundeswehr, München, Fakultät für Informatik, Workshop 13 June 1997.

[6] H. Ji, J. Hui, E. Karasan, “GoS-based pricing and resource allocation for multimedia broadband networks”, *Proceedings of the IEEE INFOCOM ’96*, San Francisco, March 24-28, 1996.

[7] L.W. Mc Knight, J. P. Bailey (editors), *Internet Economics*, MIT Press, 1997.

[8] J. K. MacKie-Mason, H. R. Varian, “Pricing Congestible Network Resources”, *IEEE Journal on Selected Areas in Communications*, 13, 1995.

[9] L. Murphy, J. Murphy, “Pricing for ATM Network Efficiency”, Proceedings of 3rd International Conference on Telecommunication Systems, Modeling and Analysis, Nashville, 1995.

[10] A. Soceanu, F. Moldoveanu, “Abrechnung im Netz”, *Gateway*, Dec. 1997.

[11] J. Walrand, P. Varaiya, *High Performance Communication Networks*, Morgan Kaufmann Publishers Inc., 1996.