Usage-based charging using effective bandwidths: studies and reality*

Vasilios A. Siris[1], David J. Songhurst[2], George D. Stamoulis[1,4], Mechthild Stoer[3]

Technical Report 243, ICS-FORTH, January 1999

Abstract

This paper studies an approach that uses the effective bandwidth as a measure of resource usage for creating time-only (flat rate) and time- and volume-based tariffs for ATM services with guaranteed QoS. We first argue that it is advantageous for a network operator to charge according to effective bandwidths since this would lead to higher aggregate utility and to competitive gains related to the long term impact of pricing. Next, we present numerical investigations involving real broadband traffic, showing how this tariffing approach can accurately and consistently take into account the effects of link and traffic contract parameters on resource usage. Finally, we compare the tariffs derived under the studied approach with real tariffs published by a particular network operator.

Keywords: tariffs, charging, competition, ATM, traffic contracts

1 Introduction

Charging, accounting and billing are crucial features of telecommunication services. How should the network provider design tariffs for the range of services offered? This is partly a marketing decision – tariffs must be attractive to customers – but network providers are also concerned with efficiency and cost-recovery. Charging schemes should encourage efficient use of the network and should generate revenue in a fair way according to the relative usage of customers.

In multiservice networks, tariffs might depend on a number of parameters defining the traffic and quality of service characteristics of a connection. The way that a customer uses the network depends on the tariffs and on how the customer values each type of connection (the customer's *utility*, in the language of economics). The challenge is to devise tariffs that are readily implementable, reflect network resource usage, and convey appropriate usage incentives to users.

*This work was supported in part by the European Commission under ACTS Project CA$hMAN (AC-039). An abridged version of this paper will appear in the 16th International Teletraffic Congress (ITC-16), Edinburgh, U.K.

[1]ICS-FORTH, P.O. Box 1385 GR 711 10 Heraklion, Greece. E-mail: {vsiris,gstamoul}@ics.forth.gr

[2]Lyndewode Research Limited, 23 Lyndewode Road, Cambridge CB1 2HN, U.K. E-mail: djs@lyndewode.co.uk

[3]Telenor AS, P.O. Box 83, N-2007 Kjeller, Norway. E-mail: Mechthild.Stoer@telenor.com

[4]Dept. of Computer Science, University of Crete, Heraklion, Greece.

The theory of effective bandwidth has proved a fruitful way to characterise resource usage in multiservice networks, and considerable attention has been given to tariffs based on simple bounds to effective bandwidth functions [12, 5]. Within the ACTS project CA$hMAN (Charging and Accounting Schemes for Multiservice ATM Networks), funded by the European Commission, a range of usage-based charging schemes were developed, implemented, and tested in trials using an advanced experimental platform.

This paper is based, in part, on some of the work done within CA$hMAN. Our aim here is to explore some of the properties of usage-based charging schemes based on effective bandwidths: to show how network operators can benefit from such tariffs in a competitive environment, to show how simple tariffs generate charges that correctly reflect resource usage, and to compare these tariffs with some currently published ATM tariffs.

Our focus is not on optimal pricing, as is the work in [10, 13]. Our approach is to use the effective bandwidth as a proxy of resource usage rather than price bandwidth and buffers separately, as done in [10]. Approaches to optimal pricing can be built on top of the effective bandwidth approach by multiplying the measure of resource usage with an appropriate factor. Finally, we do not address architectural issues such as receiver pricing and where charges are computed. However, our approach can be used in conjunction with proposals such as expected capacity [3] and edge pricing [11], which address such issues.

The rest of the paper is structured as follows. In Section 2 we review the use of the effective bandwidth as a basis for creating simple tariffs. In Section 3 we explore the incentives for network operators to introduce tariffs based on effective bandwidths. We use models where several traffic types, with different characteristics, can choose between competing operators: A short-term model assumes that network capacity is constant, and a long-term evolutionary model assumes that network capacity changes to match demand. In Section 4 we describe how tariffs can be constructed to take proper account of ATM traffic parameters, and we investigate the advantages of using charges based on both time and volume compared with time-only (flat rate) charges. In Section 5 we examine ATM tariffs published by a particular network operator (British Telecommunications) and explore how they compare with tariffs based on effective bandwidths. Finally, in Section 6 we summarise the main points of the paper and identify issues for further investigation.

2 Tariffs based on effective bandwidths

The effective bandwidth of a source that produces a load $X(t)$ in an interval t can be defined as [9]:

$$\alpha(s,t) = \frac{1}{st} \log E\left[e^{sX(t)}\right] \tag{1}$$

for particular choices of s and t. In particular, s is a *space parameter* related to statistical multiplexing capability, and t is a *time parameter* related to the time for overload to occur. The parameters s, t together represent a *network operating point*, which depends on factors such as bandwidths, buffer sizes, and traffic mix, and can be estimated from traffic measurements [7]. The effective bandwidth has the property that it increases from the mean to the peak value of $X(t)/t$ as s increases from 0 to ∞.

For any given traffic stream, the effective bandwidth definition (1) is a template that must be filled with the network operating point parameters s, t in order to provide the correct measure of effective usage. Investigations with real traffic [7] have shown that the above effective bandwidth definition can accurately quantify resource usage.

Now let $\bar{\alpha}(m, \mathbf{h})$ be an upper bound for the greatest effective bandwidth possible subject to the traffic source $X(t)$ being constrained to have mean m and to satisfy traffic contract parameters $\mathbf{h}$. This upper bound is an appropriate basis for charging, but in order to construct usable tariffs we must find suitable approximations that simplify the formula. Assume that the source is policed by leaky buckets with parameters (ρ_k, β_k) for $k \in K$, which are a part of the traffic contract $\mathbf{h}$, and let $H(t) := \min_{k \in K}\{\rho_k t + \beta_k\}$. Then a bound, which we shall call the *simple bound*, is given by

$$\bar{\alpha}(m, \mathbf{h}) = \frac{1}{st} \log \left[1 + \frac{tm}{H(t)} \left(e^{sH(t)} - 1\right)\right] . \tag{2}$$

An important property of $\bar{\alpha}(m, \mathbf{h})$ is that it is concave in m.

Next we describe a charging scheme based on the simple bound (2) which is linear in measurements of time and volume. The approach was first introduced in [8], and later extended in [6]. The user is offered tariffs corresponding to tangents to this effective bandwidth function. A tangent at the point m has the form $f(m, \mathbf{h}; M) = a(m, \mathbf{h}) + b(m, \mathbf{h})M$, where the coefficients are given by

$$b(m, \mathbf{h}) = \frac{e^{sH(t)} - 1}{s[H(t) + mt(e^{sH(t)} - 1)]}, \qquad a(m, \mathbf{h}) = \bar{\alpha}(m, \mathbf{h}) - mb(m, \mathbf{h}) . \tag{3}$$

The user is charged at rate $a(m, \mathbf{h}) + b(m, \mathbf{h})M$ per time unit, where M is his actual (measured) mean rate. This gives a charge of $a(m, \mathbf{h})T + b(m, \mathbf{h})V$ for a connection of duration T and volume V. The network may additionally charge a fixed fee $c(m, \mathbf{h})$ for each connection, which represents the cost, in switching and signaling resources, of establishing a new connection. The user then sees a charge $aT + bV + c$ arising from the selected tariff (a, b, c), comprising a duration charge, a volume charge, and a per-connection charge. We call this the *abc scheme*, previously described in [12]. Our focus is on the usage component of the charge, and henceforth we ignore the fixed charge c.

For a given traffic contract the user may be offered several choices of tariffs, corresponding to distinct tangents to the effective bandwidth bound curve. The user's choice of tariff should depend on his estimate of the mean rate of his connection. A user with a low expected mean rate should choose a tariff with small duration charge a, whereas a user with a high expected mean rate should choose a tariff with small volume charge b (see also Section 4.3). In order to minimise the expected charge the user should choose the tariff corresponding most closely to his expected mean rate, leading to a charge which is approximately proportional to the effective bandwidth bound. The user's choice of tariff thus conveys information to the network which could be used in connection acceptance control. This charging scheme thus provides appropriate incentives to the user to choose the best tariff and to constrain the duration and volume of his connection.

3 Business models

3.1 Argument for charging according to effective bandwidths

We investigate a situation which intuitively justifies why operators have the incentive to charge according to effective bandwidths, or in general, resource usage estimates used in connection acceptance control.

An operator has a link with fixed capacity C, buffer size B, and maximum overflow probability $10^{-\gamma}$. The operator's aim is to maximise the sum of the utilities of all connections sharing the link without degrading the quality γ. To ensure the quality, the operator employs connection acceptance control. One way to do this is to estimate resource usage by effective bandwidths. In this way, n_i connections (with effective bandwidth α_i) from each type i will be accepted if $\sum_i n_i \alpha_i \leq C^*$. Formally, the right hand side depends on C, B, γ, which are constants, as well as s and t, which we also assume to be constant [6, 7]. The operator does not differentiate between connections other than by their traffic contract and their mean rate, which are summarised by the effective bandwidth.

To make the example more illustrative, we assume that there are only two types of connections, both having $PCR = 3$ Mbps, but different mean rates. We assume that the operator at first sets tariffs that are based on peak rate solely, and that the charge per time unit for both connection types is $g = 10$ money units. We apply the simple bound (2) of the effective bandwidth for the case of a single leaky bucket which polices the peak rate PCR, and assume that for the above value of g the customer's response (n_1, n_2) lies on the boundary of the acceptance region (the region of traffic mixes admitted by connection acceptance control), i.e., it satisfies $\sum_i n_i \alpha_i \approx C^* = 128.9$ Mbps. Table 1 specifies such a traffic mix (for the specific mix, $s = 1.78$ Mbit^{-1} and $t = 0.4$ sec).

Type i	# of connections n_i	mean rate m_i	eff. bandwidth α_i
1	17	1	1.75
2	39	2	2.51

Table 1: A traffic mix for $PCR = 3$ Mbps, $C = 150$ Mbps, $B = 1500$ cells, and $P(\text{overflow}) \leq 10^{-7}$.

The operator may now reason as follows: if the charge for connection type 1 were decreased, and the charge for connection type 2 were increased, then more connections of type 1 become interested in using the link, while some connections of type 2 drop off. As we will see, this can theoretically be done in a way that the connection admission criterion is still satisfied, but the total utility increases. Indeed, assume a known utility $U_i(n_i)$ (in terms of "produced" quantity of connection type i), and that there are no cross-price relations between the two "products". The utility's derivative $U_i'(n_i)$ is the charge per time unit of connection type i that leads to a demand of n_i connections. For the present traffic mix $(n_1, n_2) = (17, 39)$ the operator charges g per time unit and connection type, and sells quantities (n_1, n_2), therefore $U_i'(n_i) = g$. Now, if the traffic mix (n_1, n_2) could, by adjusting prices, be changed to $(n_1 + \delta, n_2 - \delta\alpha_1/\alpha_2)$ for some small value δ, then the new traffic mix would still be admitted by acceptance control since we have moved on the boundary of the acceptance region. However, the total utility $U_1 + U_2$ would increase by

$\delta(U_1'(n_1) - \alpha_1/\alpha_2 U_2'(n_2)) = \delta g(1 - \alpha_1/\alpha_2) = 3.03\delta$. This is the operator's incentive for changing the tariff structure by charging connection type 1 less and connection type 2 more. The incentive is larger, the larger the difference in effective bandwidths. Extending this argument, it follows that optimising the utility over the acceptance region will ideally lead to a tariff structure where charges are proportional to the resource usage estimates used in connection acceptance control.[1]

3.2 Consequences of tariffs based on different resource usage estimates

We consider the long-term effects of introducing tariffs that take into account peak and mean rates, compared with keeping a tariff based on peak rates only. We illustrate these effects using an evolutionary model with two competing network operators. Each operator charges according to its estimates of resource usage costs of different connection types. We assume that, for each operator, available capacity matches demand at each stage of the evolution.

Stage A: Both operators use resource allocation and tariffs based on peak rates.

Stage B: Each operator finds from measurements that the aggregate effective bandwidth of his traffic is less than his capacity (which equals the aggregate peak rate of his traffic). Thus, more efficient resource allocation can be attained, more customers can be accepted, and charges are reduced correspondingly.

Stage C: Operator 1 introduces *abc* tariffs with acceptance control based on estimates of per connection mean rates that depend on the selected tariff. Operator 1 then offers lower tariffs for connections with low mean rates, but higher tariffs for connections with high mean rates.

Stage D: Connections with mean rates lower than the overall average have an incentive to migrate to Operator 1. Connections with higher than average mean rates have an incentive to migrate to Operator 2. As a result, Operator 2 finds that its connections have an increased overall mean rate. Operator 2 must then use more resources per connection and has to increase its charges.

Stage E: More connection types now have an incentive to migrate to Operator 1, again increasing the per connection costs for Operator 2.

Stage F: In the limit Operator 2 has to return to tariffs and resource allocation by peak rate, and only retains traffic with mean rate = peak rate. For all other traffic types Operator 1 has lower usage costs and can charge at a lower rate.

A numerical example of evolution of charges for three connection types is shown in Table 2. Stages A, B, C, D, F are as described above. The connection types C1, C2, and C3 have mean rates 7, 4, and 1, and all have peak rate 10 (all rates are in Mbps). Effective bandwidths (hence tariffs) are computed using the simple bound (2) for a single leaky bucket (with $\beta = 0$) and $st = 0.3$ sec/Mbit. At stage B the operators find, by estimating the aggregate effective bandwidth using measurements, that they can accept 39% more connections from each traffic type. From stage C onwards Operator 1 offers *abc* tariffs and

[1] One can also argue this by considering the Lagrangian optimisation problem formulation, and the associated shadow prices, of the connection acceptance control constraint [6].

each connection type chooses the tariff optimal for its mean rate. At stage D 50% of C1 connections have moved from Operator 1 to Operator 2 and 50% of C3 connections have moved from Operator 2 to Operator 1. At the limiting stage F, Operator 2 has only C1 connections, since all connections with lower mean rate have moved to Operator 1. Note that the assumption of 50% of the connections migrating to the most affordable operator does not affect the qualitative conclusions.

	Operator 1			Operator 2		
Stage	C1	C2	C3	C1	C2	C3
A	10	10	10	10	10	10
B	7.20	7.20	7.20	7.20	7.20	7.20
C	8.85	7.20	3.56	7.20	7.20	7.20
D	8.85	7.20	3.56	7.85	7.85	7.85
F	8.85	7.20	3.56	8.85	8.85	8.85

Table 2: Evolution of comparative charges for different connection types.

This evolutionary scenario illustrates the economic principle of *adverse selection*, whereby differential pricing causes customer migration which ultimately leaves one supplier (the one that charges based on peak rates only) with the worst case (highest mean) traffic.

4 ATM tariffs based on effective bandwidths

In this section we derive tariffs for ATM connections showing how our approach can accurately and consistently capture the effects of link and traffic contract parameters on the resource usage of a connection, and how it can be used to derive both time-only (flat rate) and time- and volume-based tariffs. These can be subsequently modified to take into account other factors that affect final tariffs, such as economic and marketing issues.

Parameters s, t, as discussed in Section 2, are system-defined parameters that depend on the characteristics of the multiplexed traffic and the link resources (buffer and capacity), and particular pairs will be taken to characterise periods of the day during which the traffic mix remains relatively constant. In our investigations we consider values of s, t that correspond to traffic mixes containing MPEG-1 compressed video and voice traffic.

4.1 ATM traffic contract parameters and leaky bucket parameters

First we discuss the relation of traffic contract parameters to leaky bucket parameters for Variable Bit Rate (VBR) services. Our discussion holds for both real-time VBR and non-real-time VBR services, the difference between the two being that the rt-VBR service has explicit delay requirements, whereas the nrt-VBR service does not.

Three source traffic descriptors have been defined for VBR connections [1]: the peak cell rate (PCR), the sustainable cell rate (SCR), and the maximum burst size (MBS). The compliance of an ATM connection to its traffic contract is done using the Generic Cell Rate Algorithm (GCRA), which determines for each cell arrival, whether the cell conforms to the traffic contract. The GCRA includes two parameters, T and BT (burst tolerance), and is equivalent to a continuous leaky bucket with leak rate 1 and bucket size $T + BT$, for which the contents increase by T for each conforming cell. The compliance of a VBR connection involves two GCRA's[2]: GCRA$(1/PCR, 0)$, for the compliance of the PCR, and GCRA$(1/SCR, BT)$ for the compliance of the SCR. The value of BT is computed by the traffic contract parameters using the equation $BT = (MBS-1)(1/SCR-1/PCR)$.

The leaky bucket parameters that correspond to the above conformance definition are computed as follows. For the PCR, we have the leaky bucket $(\rho_0, \beta_0) = (h, 0) = (PCR, 0)$. For the GCRA that polices the SCR, one can show that for the maximum number of conforming back-to-back cells to equal MBS, the bucket size must be $(MBS - 1)(1 - SCR/PCR) + 1$. Hence, the leaky bucket for the SCR is $(\rho_1, \beta_1) = (SCR, (MBS - 1)(1 - SCR/PCR) + 1)$.

4.2 Time-only charging

With time-only tariffs, a connection is charged based on the worst-case output that complies to its traffic contract. Hence, a connection's charge per time unit will be proportional to the upper bound of the effective bandwidth for its traffic parameters. For the remainder of this section we use the simple bound (2), where $H(t) := \min_{k=0,1}\{\rho_k t + \beta_k\}$, with ρ_k, β_k for $k = 0, 1$ computed as described above. Since we assume that a connection is charged based on the worst-case output, we take the mean rate m that appears in (2) to equal the value of SCR.

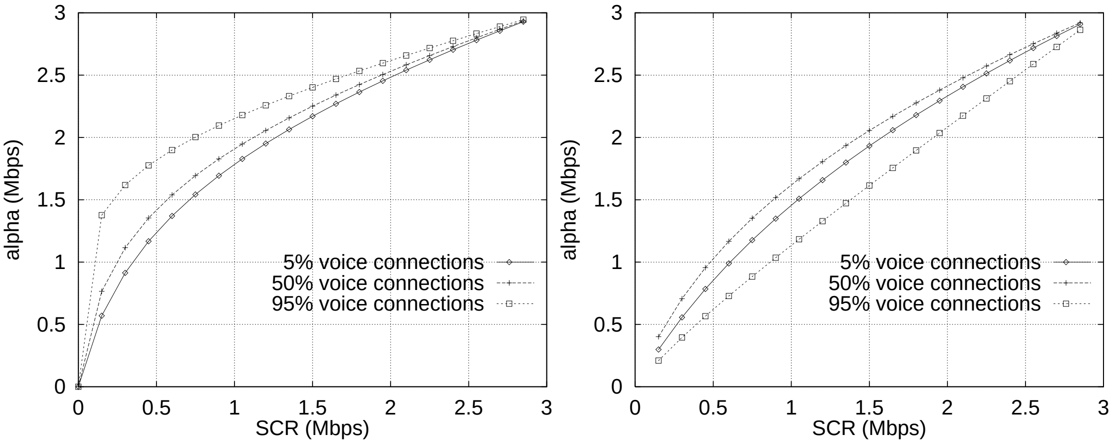
Dependence of usage charges on link resources and traffic mix

Figures 1(a) and 1(b) show, for $PCR = 3$ Mbps and $MBS = 200$ cells, the usage charge as a function of SCR for different link capacities and different traffic mixes containing MPEG-1 and voice traffic. The figures show that the dependence of the usage charge on SCR is concave and when SCR approaches zero or PCR, the effect of the traffic mix on the usage charge diminishes. Furthermore, comparing the two figures we see that for the same values of PCR, SCR, and MBS, the usage charge is lower for capacity $C = 155$ Mbps than for $C = 34$ Mbps. This is because the leaky bucket traffic characterization is tighter for higher link capacities.

Dependence of usage charges on leaky bucket parameters

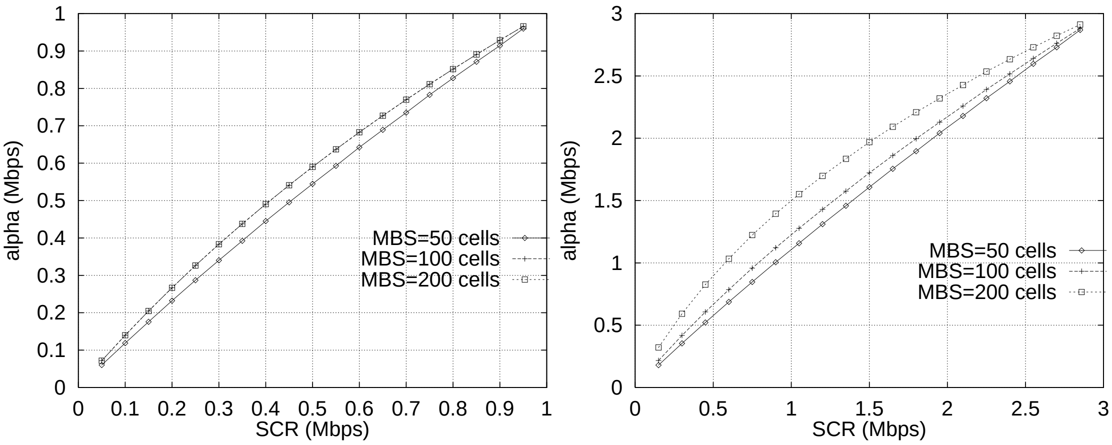
Figures 2(a) and 2(b) show, for $PCR = 1$ and 3 Mbps, the dependence of the usage charges on SCR for various values of MBS that are typical in commercial provision of ATM services (namely, $MBS = 50, 100, 200$ cells). Figure 2(a) shows that the value of MBS does not always affect the usage charge, since the curves for $MBS = 100$ and 200 cells coincide. In particular, it affects the charge only when $H(t)$ that appears in (2) is determined by the leaky bucket which polices SCR. Figure 2(b) shows that the concavity of the usage charge increases, as the value of MBS increases. Comparison of Figures 2(a) and 2(b) shows that, typically, the effect of MBS on usage is larger for larger values of PCR. Furthermore, as was the case for the dependence on the traffic mix, for values of SCR close to zero or PCR, the effect of MBS is smaller.

[2] For simplicity we do not consider the Cell Delay Variation Tolerance ($CDVT$) [1], since it will typically be much smaller than BT.



(a) $C = 34$ Mbps (b) $C = 155$ Mbps

Figure 1: Dependence of time-only usage charges on traffic mix and link capacity. $PCR = 3$ Mbps and $MBS = 200$ cells. (s, t) corresponds to a traffic mix of MPEG-1 and voice connections (each carrying 30 voice channels), multiplexed in a link with $B = 4$ msec and $P(\text{overflow}) \leq 10^{-7}$.



(a) $PCR = 1$ Mbps (b) $PCR = 3$ Mbps

Figure 2: Dependence of time-only usage charges on traffic contract parameters. (s, t) corresponds to a traffic mix of MPEG-1 and 20% voice connections, multiplexed in a link with $C = 155$ Mbps, $B = 4$ msec, and $P(\text{overflow}) \leq 10^{-7}$.

4.3 Time- and volume-based charging

In this section we present some specific examples of time- and volume-based tariffs using the approach described in Section 2, and we compare time-only with time- and volume-based charging.

m (Mbps)	a	b
0.30	0.45	1.47
0.60	0.72	0.82
0.90	0.91	0.57
1.50	1.16	0.35

(a) $PCR = 3$ Mbps, $SCR = 1.5$ Mbps, $MBS = 200$ cells

m (Mbps)	a	b
0.20	0.16	1.60
0.40	0.32	1.01
0.60	0.46	0.74
1.10	0.70	0.44

(b) $PCR = 1.5$ Mbps, $SCR = 1.1$ Mbps, $MBS = 200$ cells

Table 3: Tariffs for (s, t) corresponding to a traffic mix of MPEG-1 and 60% voice connections multiplexed in a link with $C = 155$ Mbps, $B = 4$ msec, and $P(\text{overflow}) \leq 10^{-7}$.

Tables 3(a) and 3(b) show the tariffs for two traffic contracts. As expected, due to the concavity of the effective bandwidth bound as a function of the mean rate, the duration charge a, which is the ordinate of the point where the tangent to the bound intersects the vertical axis, increases with the mean rate. On the other hand, the volume charge b, which is the slope of the tangent, decreases with the mean rate.

Once again, we can argue why time- and volume-based tariffs are advantageous compared to time-only charges. Consider two users: User 1 with traffic having mean rate $M = 0.26$ Mbps, and User 2 with traffic having mean rate $M = 1.5$ Mbps. Both users have the same traffic contract with parameters $PCR = 3$ Mbps, $SCR = 1.5$ Mbps, and $MBS = 200$ cells; Moreover, both users are assumed to be rational, thus aiming to minimise their charges. User 1 minimises his charge by selecting the tariff in Table 3(a) that corresponds to the mean rate closest to his actual mean rate, i.e., the tariff pair $(a, b) = (0.45, 1.47)$, leading to a charge per second equal to $a + b \times 0.26 = 0.83$. On the other hand, User 2's charge per second is $a + b \times 1.5 = 1.69$, attained for his optimal tariff selection, which is the pair $(a, b) = (1.16, 0.35)$ in Table 3(a). From the above, User 1's charge per second (0.83) is approximately 50% of that of User 2 (1.69). If the provider offered solely time-only tariffs, both users would incur the same charge.

In practice, it is likely that both time-only and time- and volume-based tariffs will appear. The former will be targeted to applications which send traffic very close to the maximum that is allowed by their traffic contract, whereas the latter will be targeted to applications which send traffic that is considerably lower than the maximum amount allowed by their respective contract.

5 Comparison of effective bandwidth tariffs with published tariffs of British Telecommunications

In this section, we compare tariffs derived by our approach with tariffs of the CellStream ATM service of British Telecommunications (BT). CellStream offers PVC connections within the UK and to international destinations. Charges for this service include a site connection charge and rental, and a PVC charge. The latter contains a fixed component and a per km charge, which depend on the traffic contract parameters (PCR, SCR, and MBS). Our investigation focuses on the fixed component of the PVC charge.

BT's prices for bidirectional VBR[3] connections are defined in terms of CBR prices using the multipliers shown in Tables 4(a) and 4(b) [2], which depend on the burst ratio ($= PCR/SCR$) and the MBS respectively. The price for unidirectional PVCs is 60% of the price of the bidirectional PVC with the same traffic parameters for both directions.

For example, consider a bidirectional VBR connection with $PCR = 5$ Mbps, $SCR = 1$ Mbps, and $MBS = 100$ cells. The usage price for a CBR connection with PCR equal to the SCR of the VBR connection considered (i.e., with $PCR = 1$ Mbps) is 866 £/year. Using Tables 4(a) and 4(b) we find the multipliers 1.5 for the burst ratio (= 5), and 0.9 for the MBS (= 100 cells). Hence, the price for the VBR connection is $866 \times 1.5 \times 0.9 = 1169.1$ £/year. BT's pricing structure, in addition to Tables 4(a) and 4(b), includes one more condition: for burst ratios less than 1.8 all three MBS options are priced as $MBS = 200$ cells.

Burst ratio	1	2	5	10	15	20
Multiplier	0.9	1.1	1.5	2.0	2.3	2.5

(a) Multipliers for the burst ratio PCR/SCR

MBS (in cells)	50	100	200
Multiplier	0.85	0.9	1.0

(b) Multipliers for MBS

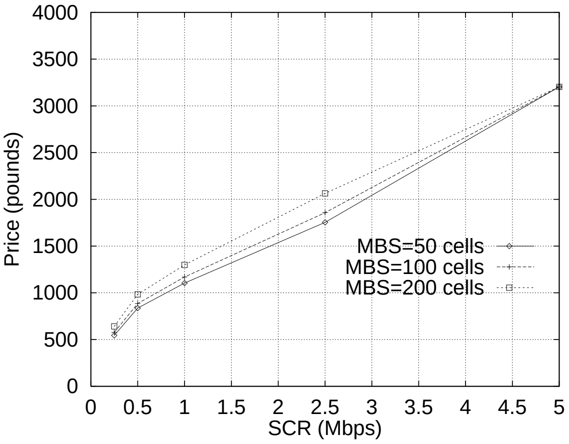
Table 4: BT price multipliers.

A first observation of BT's tariffs relates to the relative prices of CBR and VBR connections. For $PCR/SCR = 1$ and $MBS = 200$ cells, BT's VBR tariffs would use the multipliers (see Tables 4(a) and 4(b)) 0.9 and 1. Hence, a user with constant rate traffic that selects a VBR connection with $PCR/SCR = 1$ and $MBS = 200$ cells would be charged 90% of the price for a CBR connection with the same PCR. This can be justified only if there is some additional quality (e.g., priority) associated with the CBR service.

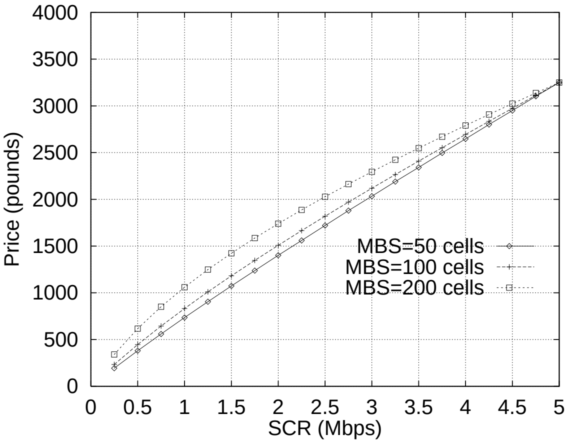
Figure 3 compares the prices using the effective bandwidth approach with BT prices, for $PCR = 5$ Mbps. In order to make the comparison, we have multiplied the effective bandwidth bound with the scaling factor 650 £/Mbps. This value was chosen so that for $SCR = PCR$ all curves yield the same charge. Figure 3(a) shows that the price given by BT for a connection with $PCR = 5$ Mbps increases quickly for small values of SCR

[3] BT also defines a service, called VBR+, where the cells in excess of the traffic contract are marked, and dropped only if there is congestion. The prices for VBR+ are also defined in terms of the CBR prices, however the burst size multipliers are typically higher than those in Table 4(a).

(less than 0.5 Mbps), and is almost linear in SCR up to $SCR = 5/2 = 2.5$ Mbps. Above that, prices for all values of MBS are equal to those for $MBS = 200$ cells. As a result, for $MBS = 50$ and 100 cells, and $SCR \geq 1$ Mbps, prices are slightly convex in SCR. For $SCR \geq 2$ Mbps these curves were obtained by means of linear interpolation, because there is no data in the tariff tables of BT for burst ratios between 1 and 1.8. On the other hand (Figure 3(b)), the price with the effective bandwidth approach is concave in SCR. Furthermore, the concavity increases for higher values of MBS.



(a) BT prices

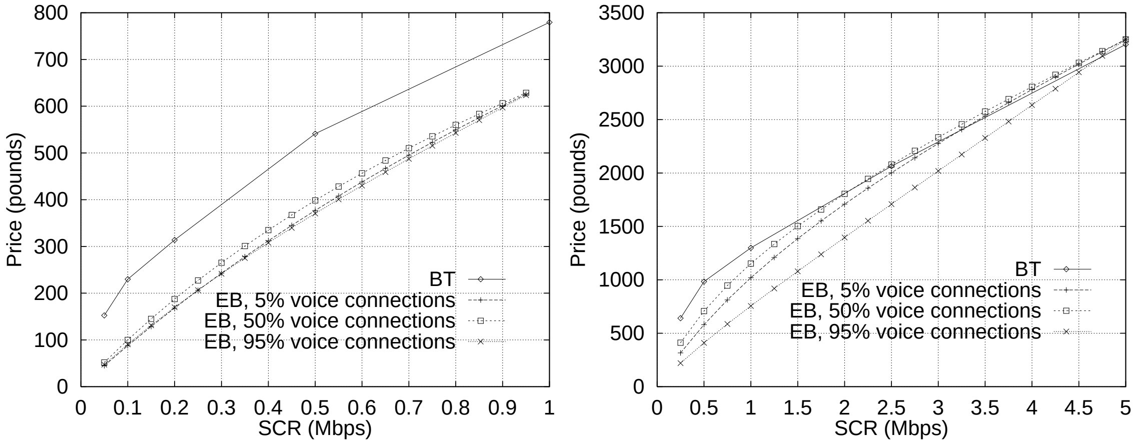


(b) Effective bandwidth prices

Figure 3: Effective bandwidth and BT prices for different MBS. $PCR = 5$ Mbps. (s,t) corresponds to a traffic mix of MPEG-1 and 20% voice connections multiplexed in a link with $C = 155$ Mbps, $B = 4$ msec, and $P(\text{overflow}) \leq 10^{-7}$.

Figures 4(a) and 4(b) compare BT prices with prices computed using our approach, for $PCR = 1$ and 5 Mbps, and $MBS = 200$ cells. Conversion of effective bandwidth into money was done using the same scaling factor 650 £/Mbps, as explained above. Figure 4(b) shows that, compared to BT prices, effective bandwidth based prices decrease faster as SCR decreases, particularly for small values of SCR. Furthermore, comparison of Figures 4(a) and 4(b) shows that, compared to BT prices, effective bandwidth based prices decrease faster with decreasing PCR. This is also shown in Figure 5 for values of PCR smaller than 2 Mbps. Figure 5 also shows that for $PCR > 6$ Mbps, prices based on the effective bandwidth approach increase more slowly compared to BT prices. These deviations suggest that BT prices are not fair, in the sense that they do not accurately reflect resource usage and can hence give the wrong incentives for resource usage. These deviations may reflect the provider's marketing strategy for orienting each market segment (i.e., traffic type) towards a certain subset of services offered (ATM, frame relay, leased lines, etc.). A similar study was carried out in [4], where published tariffs of AT&T are compared with effective bandwidth tariffs; the conclusions reported therein are similar to the above.

We end this section by noting that defining tariffs as a simple function of one tariff table (CBR tariffs in the case of BT's CellStream service), with parameters depending on the traffic contract parameters (PCR, SCR, and MBS) has the advantage of compactness



(a) $PCR = 1$ Mbps (b) $PCR = 5$ Mbps

Figure 4: Effective bandwidth and BT prices for $PCR = 1$ and 5 Mbps, and $MBS = 200$ cells. (s, t) corresponds to a traffic mix of MPEG-1 and voice traffic, with 5%, 50%, and 95% voice connections multiplexed in a link with $C = 155$ Mbps, $B = 4$ msec and $P(\text{overflow}) \leq 10^{-7}$.

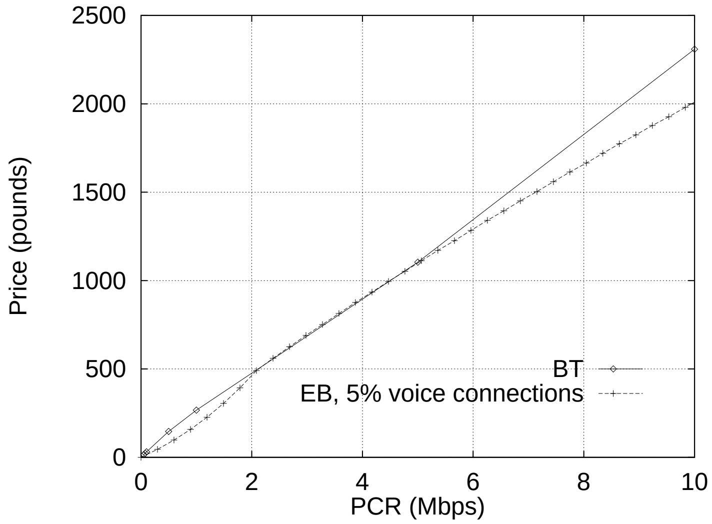


Figure 5: Effective bandwidth and BT prices for $PCR/SCR = 5$ and $MBS = 200$ cells. (s, t) corresponds to a traffic mix containing MPEG-1 and 20% voice connections multiplexed in a link with $C = 155$ Mbps, $B = 4$ msec, and $P(\text{overflow}) \leq 10^{-7}$.

and simplicity. An interesting extension would be to define tariffs produced using the effective bandwidth approach in such a way.

6 Conclusions

We have studied an approach for usage-based charging of ATM services with guaranteed QoS. The approach uses the effective bandwidth as a measure of resource usage for creating time-only (flat rate) and time- and volume-based tariffs. We have argued that it is advantageous for a network operator to employ tariffs based on effective bandwidths, because this would lead to both higher aggregate utility and to competitive gains related to the long term impact of charging.

We have presented numerical investigations involving real traffic, showing how the tariff approach studied can take into account the effects of link and traffic contract parameters on resource usage both accurately and consistently. Finally, we have compared the effective bandwidth tariffs with tariffs published by British Telecommunications, and we have noticed both similarities and discrepancies, some of which may come as a result of the provider's marketing strategy. Motivated by this, we believe that an interesting and challenging direction for future research is to extend the framework for usage-based charging to cover switched virtual channel (SVC) connections and services that are supplementary and/or complementary to ATM, taking into account the cross-elasticities in user demand for the various services as well as the provider's strategy for market segmentation.

Acknowledgements

The authors are particularly grateful to Ragnar Andreassen, Costas Courcoubetis, Frank P. Kelly, and Richard Weber for helpful discussions on the subject of this paper.

References

[1] The ATM Forum. *Traffic Management Specification Version 4.0*, 1996.

[2] BT Price List for the CellStream Service.
URL: http://www.serviceview.bt.com/list/current/docs/Cellstream/sectoc.htm

[3] D. D. Clark. Internet cost allocation and pricing. In J. P. Bailey and L. Mcknight, editors, *Internet Economics*, MIT Press, MA, 1997.

[4] CA$hMAN Consortium. *Report on User Behaviour.* Deliverable D15, 1998.

[5] C. Courcoubetis, F. P. Kelly, V. A. Siris, and R. Weber. A study of simple usage-based charging schemes for broadband networks. In *Proc. of IFIP International Conference on Broadband Communications (BC'98)*, Stuttgart, Germany, April 1998.

[6] C. Courcoubetis, F. P. Kelly, and R. Weber. Measurement-based charging in communications networks. Technical Report 1997-19, Statistical Laboratory, University of Cambridge, 1997. To appear in Operations Research.

[7] C. Courcoubetis, V. A. Siris, and G. D. Stamoulis. Application of the many sources asymptotic and effective bandwidths to traffic engineering. To appear in Telecommunication Systems. A shorter version appeared in *Proc. of ACM SIGMETRICS'98/PERFORMANCE'98*, Madison, Wisconsin, June 1998.

[8] F. P. Kelly. On tariffs, policing and admission control for multiservice networks. *Operations Research Letters*, 15:1–9, 1994.

[9] F. P. Kelly. Notes on effective bandwidths. In F. P. Kelly, S. Zachary, and I. Zeidins, editors, *Stochastic Networks: Theory and Applications*, pp 141–168. Oxford University Press, 1996.

[10] S. H. Low and P. P. Varaiya. A new approach to service provisioning in ATM networks. *IEEE/ACM Trans. on Networking*, 1(3):547–553, October 1993.

[11] S. Shenker, D. Clark, D. Estrin, and S. Herzog. Pricing in computer networks: Reshaping the research agenda. *ACM Computer Commun. Review*, pp 19–43, 1996.

[12] D. J. Songhurst and F. P. Kelly. Charging schemes for multiservice networks. In *Proc. of the 15th Intl. Teletraffic Cong. (ITC-15)*, pp 781 – 790, 1997. Elsevier Science B.V.

[13] Q. Wang, J. M. Peha, and M. A. Sirbu. The design of an optimal pricing scheme for ATM integrated-services networks. In J. P. Bailey and L. Mcknight, editors, *Internet Economics*, MIT Press, MA, 1997.