University of Würzburg

Institute of Computer Science

Research Report Series

Resource-Based Charging of ATM Connections

Norbert Vicari

Report No. 170 April 1997

Institute of Computer Science, University of Würzburg
Am Hubland, D - 97074 Würzburg, Germany
Tel.: +49 931 888 5505, Fax.: +49 931 888 4601
e-mail: vicari@informatik.uni-wuerzburg.de

Abstract

In this paper we present a resource-based charging scheme for ATM connections. The scheme is based on the application of cost functions to the resources, bandwidth and buffer, required by a VBR-ATM connection. The resource requirement is modeled with the GCRA algorithm which is used for the usage parameter control in ATM networks. Applying the charging scheme to numerical examples we demonstrate how the user behavior can be influenced by the proper choice of the parameters of the charging scheme.

1 Introduction

ATM technology is nowadays mainly used in scientific research networks or testbeds for future multiservice broadband networks. The costs for customers in those networks are of secondary interest, since in common a flat rate for the connection to the network is raised but the actual network and resource usage is not billed. For the upcoming commercial operation of broadband multiservice networks proper billing of connections is a crucial requirement to enable operators to maximize their profit and treating the customers in a fair manner. Furthermore billing schemes give the operator a measure to avoid congestion. We propose a charging scheme for VBR connections, which is oriented on the resource requirement defined by the ATM-relevant standards.

With the commercialization of the Internet different charging schemes for packet-switched networks have been proposed. MacKie [2] proposes a ‘smart market-pricing’ scheme, where each user submits a bid for each packet to submit. The network determines a cut-off price, where demand meets capacity, and transmits all packets whose bid exceed this price. In the paper of Parris [3] additionally the effects of set-up pricing (additional to the price per packet a fee for connection set-up is charged) and peak load pricing (higher charges during busy hours) are studied, assuming a wealth distribution of customers. Charging connections with respect to their priority depending on the service class is introduced in the paper [1]. Cocchi shows that under the assumption of four service classes pricing schemes can be defined that optimize network usage in the sense of a Nash equilibrium. This Internet charging schemes take not into account the dynamics of ATM traffic with the necessity of distinguishing between different burst characteristics of traffic.

More recent publications [4,5] take into account broadband multiservice traffic characteristics, but the idea of using an iterative, and therefore slow procedure to determine the price and bandwidth of connections seems not to be very promising. In [6] Murphy proposes an iterative scheme as feedback control algorithm to find an equilibrium of bandwidth demand and resources. Parris [3] introduces a pricing scheme based on the amount of bandwidth, buffer space, CPU and delay resources reserved for a connection. An approach derived from user characterization and admission control is proposed by Kelly [7]. The user is charged in accordance to the effective bandwidth and the paper proves that the charge is minimal if the user tells the truth about his bandwidth usage.

According to the ATM-Forum [8] the resource allocation schemes should make use of the *Connection Traffic Descriptor* which contains as a subset the *Source Traffic Descriptor*. The *Source Traffic Descriptor* is a set of traffic parameters which are signaled at connection setup to characterize the traffic characteristics of a particular source. The transmitted parameters depend on the class of the connection. In the proposed tariff scheme the charge for ATM connections should be related to the resources defined in the traffic contract negotiated between the user and network provider on connection setup. In Section 2 we will discuss this parameters and derive the bandwidth and buffer requirements based on a simple model. In Section 3 general principles for cost functions are introduced and applied in Section 4 to numerical examples. Section 5 concludes the paper.

2 Bandwidth and Buffer Requirement

In this section we will discuss on which class of connections the proposed charging scheme is applicable. Then the model for the derivation of the resource requirement and its parameters are presented. At the end of this Section we show an example for the resource requirement of a given connection (cell stream), which will be used for the demonstration of cost functions.

2.1 Service Categories

The ATM Forum [9] defines 5 service categories: *Constant Bit Rate* (CBR), *Real-Time Variable Bit Rate* (rt-VBR), *Non-Real-Time Variable Bit Rate* (nrt-VBR*), Unspecified Bit Rate* (UBR) and *Available Bit Rate* (ABR).

For UBR connections no *Quality of Service* (QoS) is guaranteed and for ABR connections the QoS is characterized with low but not specified *Cell Loss Rate* (CLR). Due to the nature of this kind of service only minimal resources are reserved at the connection setup and additional resources may be dynamically captured during the continuance of the connection. Charging issues of this connection type are investigated in [10].

For CBR connections the PCR and QoS are specified while for VBR connections the *Peak Cell Rate* (PCR), *Sustainable Cell Rate* (SCR), *Maximum Burst Size* (MBS) and QoS are specified on connection setup. In order to guarantee a certain QoS CBR connections require bandwidth and buffer (so called *Cell Delay Variation Tolerance* (CDVT)) to absorb the *Cell Delay Variation* (CDV) [11]. VBR connections require bandwidth and buffer, with buffer-size depending on the MBS and the CDVT. We concentrate our investigation on nrt-VBR connections, since a wider range of buffer and bandwidth combinations can be chosen. From the charging point of view the CBR and rt-VBR service categories can be considered as a special case of the nrt-VBR service category with tighter delay bounds.

2.2 nrt-VBR Service Category

For a given VBR connection the PCR, SCR, MBS and CLR are signaled in the *Source Traffic Descriptor*. Based upon this parameters the traffic contract between the user and network provider is negotiated and fixed. The conformance of cells of a connection is defined via the GCRA algorithm. The network estimates the CDV that the connection will suffer and computes the parameters for the conformance test of a connection. At the UNI the PCR and SCR will be controlled with the following parameter tests:

PCR conformance: $GCRA(\frac{1}{PCR}, CDVT)$,

SCR conformance: $GCRA(\frac{1}{SCR}, BT + CDVT)$.

The *Burst Tolerance* (BT) is computed in dependence of the PCR and SCR. For VBR connections the required CDVT is magnitudes smaller than the BT. Therefore, we assume that for charging purposes the resources bandwidth and memory can be described by SCR and BT.

2.2.1 Model Description

We model the dependency of bandwidth and buffer using the GCRA algorithm. In the model used for this investigation the following assumptions are made:

- resource requirements are described by the SCR and BT, CVD is neglected since the buffer requirement for the BT is of magnitudes larger,
- the cell arrival process is characterized by generally and independently distributed inter-arrival times with a mean inter-arrival time IAT and a coefficient of variance CoV,
- the PCR is the full link capacity.

The GCRA is analyzed as $GI^{[x]}/D/1 - s$ discrete-time queueing system [12]. In order to compare different pairs (SCR, BT) we use the ratio of non-conforming cells.

2.2.2 Example and Data Set

In Figure 1 the minimal required BT for negative binomially distributed inter-arrival times (with an IAT of 80 and CoV of 0.5 and IAT of 20 and CoV of 0.5, 1 and 2) in dependence of the SCR is depicted. The minimal ratio of non-conforming cells is set to 10^{-5}. Reducing the required bandwidth (increasing SCR) causes an exponential increase of the required buffer (BT). The SCR has an upper bound with the mean inter-arrival time IAT, where the required BT grows infinite. With increasing coefficient of variation CoV of the inter-arrival time of the cell stream, which causes a more bursty behavior of the cell stream, a growth of the required BT is observed.

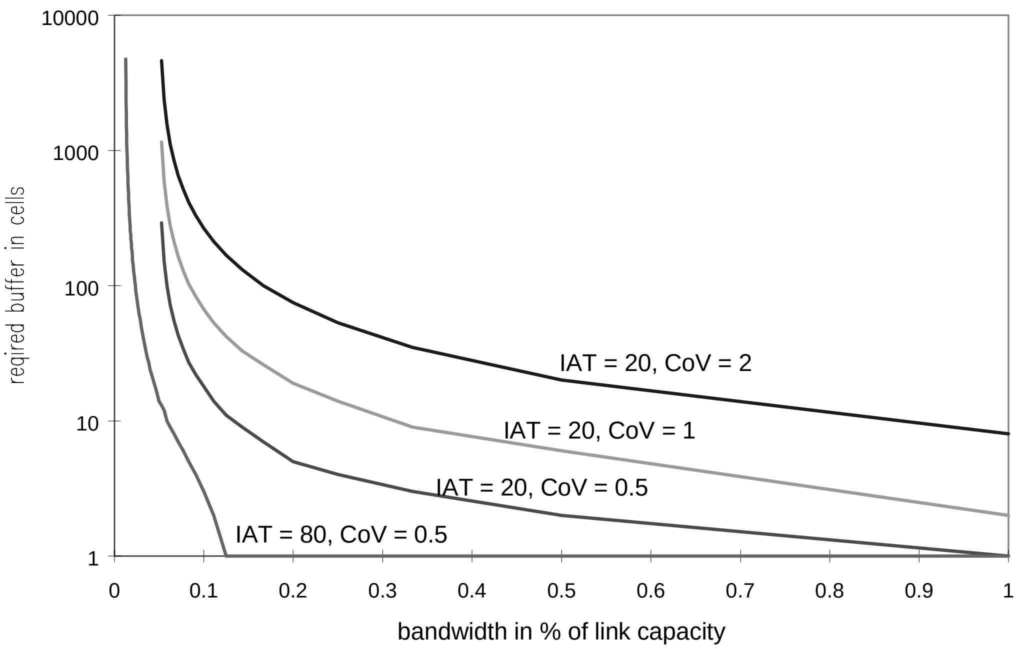

Figure 1: Minimal BT in dependence of SCR

In the next section we will first discuss different principles for cost functions and then apply those principles to the above shown resource requirement functions.

3 Cost Functions

As demonstrated in the last section, each connection requires the bandwidth 1/SCR and buffer BT. Applying a QoS criteria the buffer requirements can be computed in dependence of the SCR. Therefore, the charge c for a connection can be written as sum of the cost $c_1(SCR)$ for the bandwidth and the cost $c_2(BT)$ of the buffer.

$$c(SCR, BT) = c_1(SCR) + c_2(BT) \tag{1}$$

3.1 General Properties of Cost Functions

The cost functions for the resources buffer and bandwidth have to fulfill certain properties. In this section we will denote the ratio of the required resource with x and the general cost function with $cf(x)$. Using this properties we reduce the number of potential cost functions and their parameters.

The charge for the usage of resources is positive:

$$cf(x) > 0, \quad \text{for } x > 0. \tag{2}$$

The more resources a user requires, the higher is the charge for the corresponding connection. This implies that the first partial of the cost functions is positive:

$$\frac{\partial cf(x)}{\partial x} > 0, \quad \text{for } x > 0. \tag{3}$$

The users should get some discount for buying a larger amount of bandwidth or buffers. Therefore the second partial of the cost functions is negative:

$$\frac{\partial^2 cf(x)}{\partial x^2} < 0, \quad \text{for } x > 0. \tag{4}$$

To enforce a further reduction of the possible parameters we give limits for the maximum and minimum gradient of the cost functions. We define the cost functions on a parameter space reaching from 0.001 to 1. A cost function fulfilling the equations (2 - 4) reaches the maximum gradient at the minimum of the parameter space and the minimum gradient at the maximum of the parameter space.

$$\text{maximum gradient: } g_{\max} = \frac{\partial cf(x)}{\partial x}, \quad \text{for } x = 1. \tag{5}$$

$$\text{minimum gradient: } g_{\min} = \frac{\partial cf(x)}{\partial x}, \quad \text{for } x = 0.001. \tag{6}$$

For the description of the characteristics of the cost functions we use the ratio of the maximum and minimum gradient:

$$\text{gradient ratio: } g_{ratio} = \frac{g_{\max}}{g_{\min}}. \tag{7}$$

3.2 Cost Function Classes

We have identified 3 function classes which fulfill the conditions defined in equations (2 - 4). Type A is are root functions, type B are logarithmic functions and type C are negative-exponential functions. In the following tabular the functions are defined and the valid parameters are identified.

type	**function**	**valid parameters**
type A	$a \cdot x^b$	$a > 0, \quad 0 < b < 1$
type B	$a \cdot \ln(bx)$	$a > 0, \quad b > \max_x \frac{1}{x}$
type C	$a \cdot (1 - e^{-bx})$	$a > 0, \quad b > 0$

3.3 Parameter Discussion

For the comparison of the different types of functions the parameters are chosen to obtain similar properties (in terms of identical ratio of maximum and minimum gradient and identical minimum gradient) for the functions. The maximal costs are set to 1.

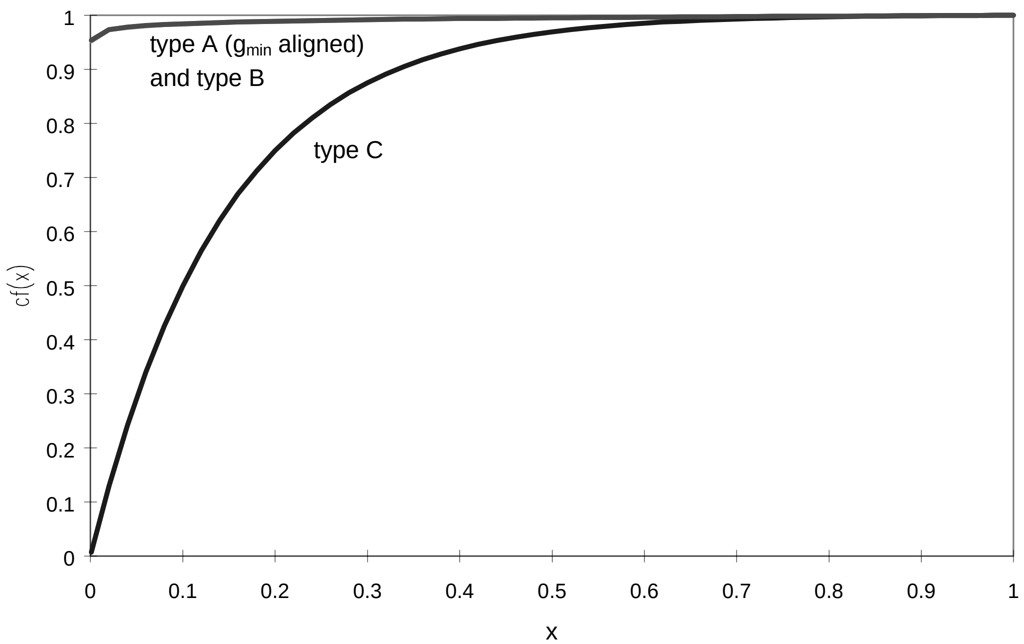

Figure 2: Comparison of function types A, B and C for g_{ratio} = 1000.

Since the relation of maximum and minimum gradient is constant (g_{ratio} = 1000) for function type B, the parameters of function type C are derived for this gradient ratio. This function is depicted in Figure 2 in comparison to the function types A and B with parameters derived to fit the minimum gradient of function type C. For this set of parameters only function type C covers the full cost range, whereas the functions of type A and B take only cost values in the range from 95% to 100%.

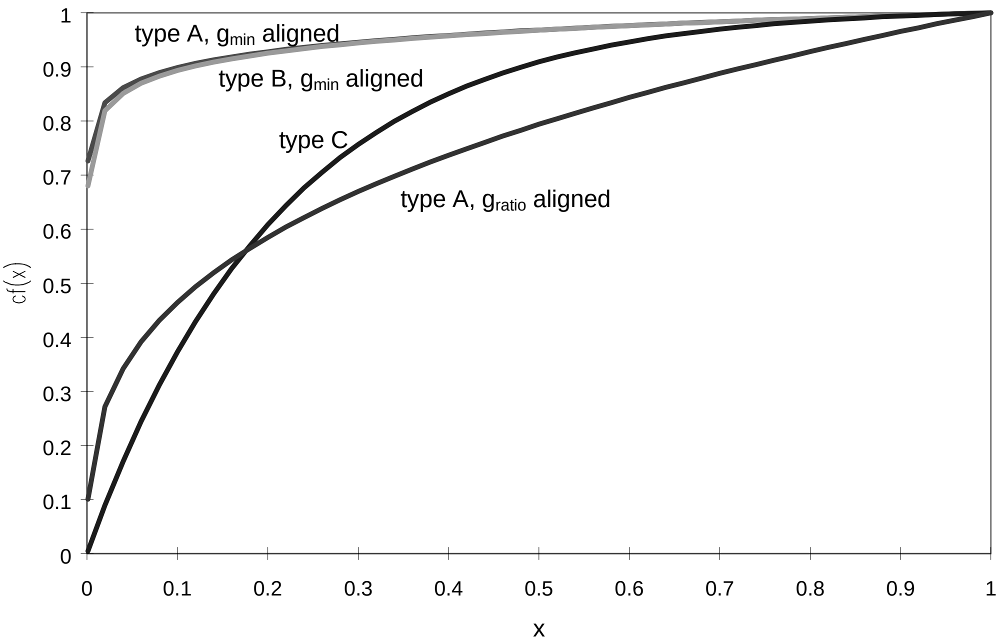

Figure 3: Comparison of function types A, B and C for g_{ratio} = 100.

In Figure 3 the gradient ratio of function type C is set to 100. The functions of type A and B are compared to this function. Again function type C covers the whole range of possible cost values. For function type A and B the parameters are derived so that the minimum gradient is aligned with the minimum gradient of function type C. The curves are nearly identical and cover only the upper 30% of the possible function values. Aligning the values of g_{ratio} of the function types A and C the minimum gradient of function type A is larger then the minimum gradient of function type C.

As reference curve we depict in Figure 4 the function type C with g_{ratio} equal 10. This curve is compared with the functions for type A and B. For function type B it is not possible to align the minimum gradient to that of function type C within the valid parameter range, causing negative function values for small values of x. The function of type A with minimum gradient alignment produces higher values than function type C, which is a result of the higher gradient ratio. Comparing the functions of type A and C with g_{ratio} equal 10 the function of type A has a greater minimum gradient as the function of type C, which alters the gradient slower.

To sum up the parameter discussion, the logarithmic (type B) functions have the disadvantage that the ratio of minimum and maximum gradient is constant, causing problems in the selection of appropriate parameters. For both, the root (type A) and negative-exponential (type C) functions, the ratio of the maximum and minimum gradients is variable. The gradient of the type A functions changes rapidly in the first third of the parameter space, whereas the gradient of the type C functions changes smoothly over the whole parameter space. Since those smooth changes are preferable for charging functions we will concentrate in the following Section on the negative-exponential functions.

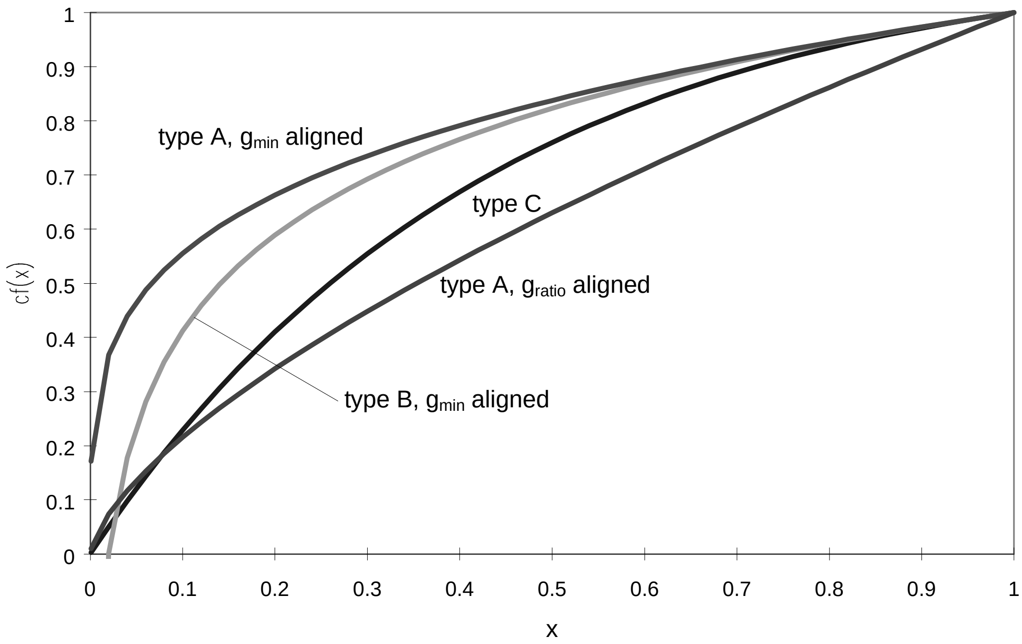

Figure 4: Comparison of function types A, B and C for a gradient ratio of 10.

4 Application of Cost Function

In this Section the influence of the choice of parameters and weight of the cost functions for the components bandwidth and buffer of a connection is investigated. The effect of combining the requirements of multiple connections to one virtual path is demonstrated in the last subsection of this part.

4.1 Influence of the Gradient Ratio

To evaluate the effect of different parameters of the cost functions for the required buffer we assign the parameters of the cost function of the bandwidth to achieve a gradient ratio of 10. The values of g_{ratio} of the functions for the charge of the required buffer are set to 10, 100 and 1000. In order to apply the cost functions, which are defined for a parameter space between 0 and 1, we normalize the required buffer under the assumption that the maximum buffer per connection is limited to 10000 cells.

In Figure 5 the curves for the total charge of the required resources of a connection with a mean cell inter-arrival time of 20 and a coefficient of variation of 2 are showed with bold lines and the curves for a cell stream with CoV 0.5 of the inter-arrival times are displayed in thin lines. All shown cost functions have a bandwidth and buffer combination where the charge for the connection is minimal and thus will be most probably chosen by the user. With higher g_{ratio} values the cost function for the buffer reacts more sensitive to a increase of the buffer requirement. In addition to the costs for the bandwidth this fact leads to a shift of the minimum costs towards a higher bandwidth. The effect can be observed for both example connections.

Of minor importance is the influence of the different cost functions of the buffer to the absolute value of the minimal charge for the connection, since the charging scheme has to be consistent for itself.

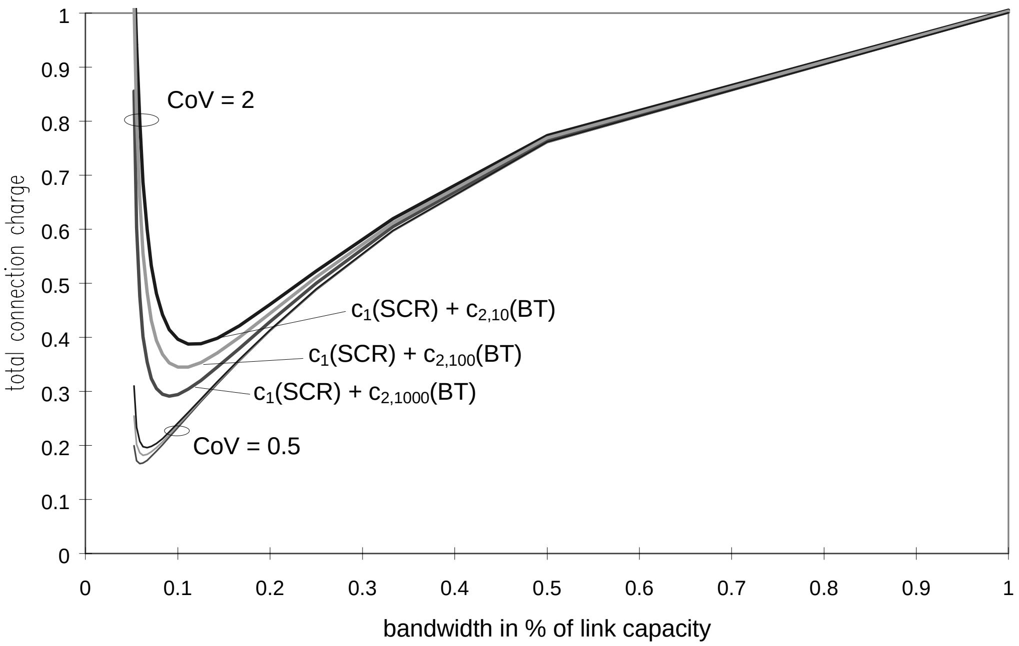

Figure 5: Influence of varying gradient ratios of the buffer cost function

4.2 Influence of Scaling

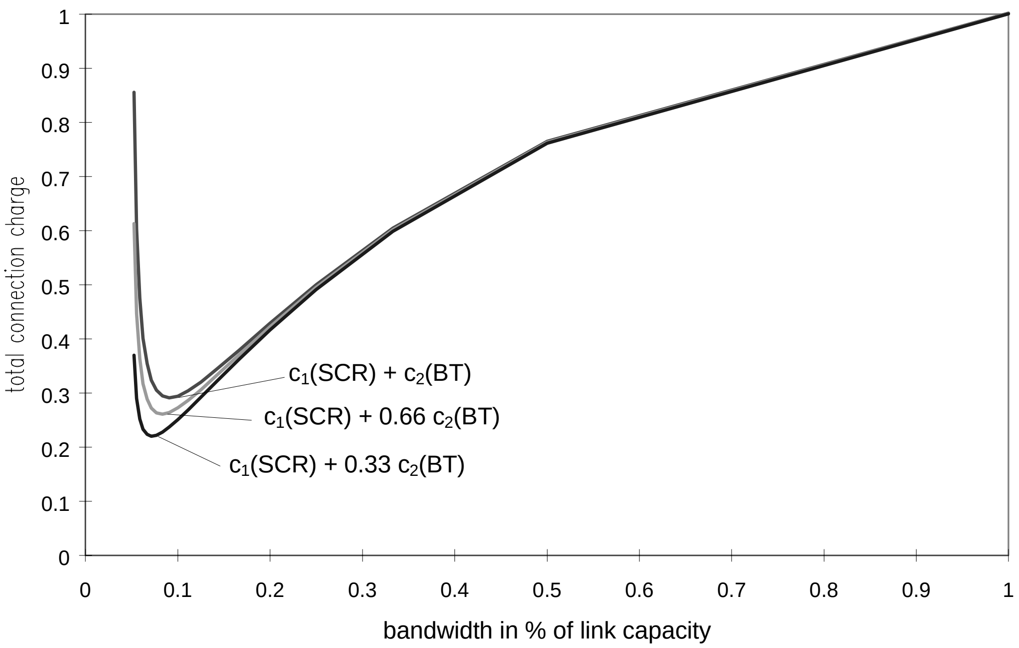

Figure 6: Influence of different weights of the buffer cost function

The influence of the charge for the required buffer of a connection can be changed introducing a weight factor w. The total charge of a connection is derived with the following formula:

$$c(SCR, BT) = c_1(SCR) + w \cdot c_2(BT) \quad (8)$$

The curves in Figure 6 are based on the previous data set, with resources derived for a cell stream with negative binomial distributed inter-arrival times with IAT 20 and CoV 2. The parameters of the component cost functions are set to achieve a gradient ratio of 10. Reducing the weight of the charge of the required buffer from 1 to 0.66 and 0.33 the minimum of the total charge is reached at smaller bandwidth in comparison to the non-weighted cost function. As a side effect of weighting the buffer cost we obtain a reduction of the minimum charge for the connection.

4.3 Discount for High Resource Usage

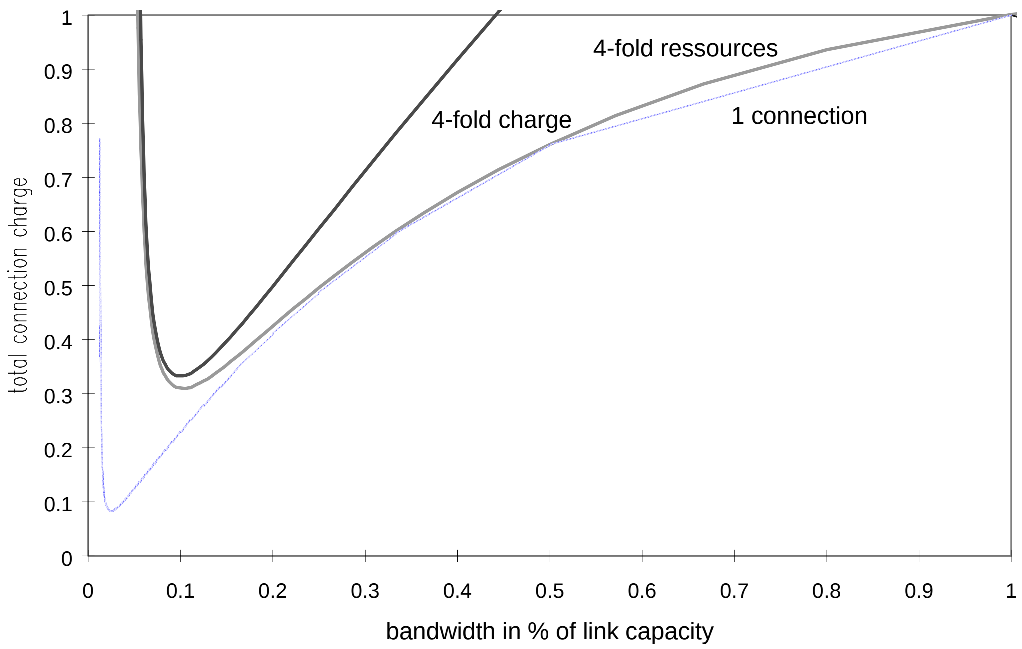

Figure 7: Discount effect of the proposed charging scheme

The properties of the component cost functions are chosen in a way that a customer is given a discount for higher resource usage. To demonstrate this behavior of the charging scheme we compare the 4-fold charge for a single connection with the charge for the 4-fold resources if they are used by a single connection.

In Figure 7 we apply for both components, bandwidth and buffer, cost functions with a gradient ratio of 10 and add them equally weighted. The resource requirement is derived from a cell stream with negative binomial distributed inter-arrival times with an IAT of 80 and a CoV of 0.5. The charge for this connection is displayed with the thin line. For four connections of this type the charge is 4 times higher, which is illustrated with the dashed line. The solid line shows the charge for a single connection with the same resource requirement. In comparison to that the minimum charge for a single connection with 4-fold resource requirement is 10% lower.

5 Conclusion

A resource based charging scheme for VBR-ATM connections was presented. The scheme consists of different cost functions which are applied to the required resources of a connection. We assume that a connection requires resources in terms of bandwidth and buffer.

To determine the dependence of bandwidth and buffer we use a model based on the GCRA algorithm. For this purpose the connection is described as cell stream with generally and independently distributed inter-arrival times. The pairs of bandwidth and buffer which cause the equal cell loss rate are considered to describe the buffer requirement in dependency of bandwidth.

With respect to general properties of cost functions the root, logarithmic and negative-exponential function types are identified as possible cost functions for the resource components. By restricting the values of this functions to the interval between 0 and 1 and exploiting the gradient ratio of this functions we obtain a measure for comparison. The negative-exponential function type is most promising, since it allows an arbitrary choice of the gradient ratio without limiting the range of the values. Further this function type has the advantage to change the gradient slowly over the parameter space, which enables a wide application range.

In the last part of the paper the identified function type with different parameters is applied to example data sets of bandwidth and buffer. For all parameters the total charge of the connection reaches a minimum. This minimum is shifted to favor higher bandwidth and lower buffer requirement if the gradient ratio of the cost function for the buffer is increased. On the other hand the minimum is shifted towards lower bandwidth and higher buffer requirement if the cost function for the buffer is less weighted.

The properties of the cost functions are preserved in their application to the required resources. This is shown by an example in which the costs of multiple connections are related to the charge for identical resources in a single connection.

The resource-based charging scheme is easy to understand and therefore can count on acceptance by customers. On the other hand the charging scheme gives the network provider a means to influence the user behavior. By a proper choice of the minimum of the cost functions a uniform usage of all resources and therefore a high network load can be achieved.

6 Acknowledgment

The author highly appreciates the financial support of the Deutsche Telekom AG (Technologiezentrum Darmstadt).

7 References

1 R. Cocchi, S. Shenker, D. Estrin and L. Zhang. *Pricing in Computer Networks: Motivation, Formulation and Example*. IEEE/ACM Transactions on Networking, Vol. 1, pp. 614 - 627, December 1993.

2 J. K. MacKie-Mason and H. R. Varian. *Pricing the Internet*. Second International Conference on Telecommunication Systems, Modeling and Analysis, Nashville, TN, March 1994, pp. 178-393.

3 C. Parris, S. Keshav and D. Ferrari. *A Framework for the Study of Pricing in Integrated Networks*. Technical Report TR-92-016, International Computer Science Institute, Berkeley, CA, March 1992.

4 S. H. Low and P. Varaiya. *A New Approach to Service Provisioning in ATM Networks*. IEE/ACM Transactions on Networking, Vol. 1, pp. 547 - 553, October 1993.

5 J. Murphy, L. Murphy and E. C. Posner. *Distributed Pricing for Embedded ATM Networks.* Proceedings of the 14th International Teletraffic Congress, Antibes, France, June 1994, pp. 1053 - 1063.

6 L. Murphy and J. Murphy. *Pricing for ATM Network Efficiency*. Third International Conference on Telecommunication Systems, Modeling and Analysis, Nashville, TN, March 1995, pp. 349-356.

7 F. P. Kelly. *On tariffs, policing and admission control for multiservice networks*. Operations Research Letter, Vol. 15, 1994.

8 The ATM Forum Technical Committee. *ATM User-Network Interface Specification, Version 3.1*. September 1994.

9 The ATM Forum Technical Committee. *Traffic Management Specification, Version 4.0.* April 1996.

10 C. Courcoubetis, V. A. Siris and D. Stamoulis. *Integration of Pricing and Flow Control for Available Bit Rate Services in ATM Networks*. Proceedings IEEE Globecom'96, London, UK, November, 1996.

11 J. Roberts, U. Mocci and J. Virtamo (Eds.). *Broadband Network Teletraffic, Final Report of Action COST 242*. Springer, 1996.

12 F. Hübner. *Dimensioning of a Peak Cell Rate Monitor Algorithm Using Discrete-Time Analysis*. Proceedings of the 14th ITC, Antibes, 1994, pp. 1415 - 1424.