form AAL2-CP packets without further processing (no SSCS).

However, many applications either generate data in a way that AAL2-CP can not be directly used (i.e. packets longer than 45 octets) or require services not provided by the AAL2-CP (like error protected payload or protection against packet loss). In these cases an SSCS is needed as interface between the user of the AAL2 (application) and the AAL2-CP.

At this moment, work is undergoing to define two SSCSes, one for the transmission of long packets and one for the transport of trunking services (land-voice). For the first, the specification is already frozen as ITU-T I.366.1 and ready for definitive approval next year. For the second, work is very intensive at this moment and the target is to have a frozen specification during 1998.

5. Efficiency of AAL2

It has been shown that AAL2 provides good mechanisms for fine adjustment of packetization delay and for achieving high BW efficiency for low bit rate and VBR application. The cost is some overhead in the form of headers. This overhead is not important at all when comparing to techniques as partial filling of cells (AAL2 behaves at least as well as partial filling) but may be significant when users have a choice among several AALs. For example, for transmitting long packets (longer than 45 octets) the user may choose either AAL2 (through its segmentation SSCS) or AAL5 (explicitly specified for long packet data). It is known that for small packets (up to a few hundred octets) AAL2 is more efficient than AAL5 while for long packets AAL5 is definitely recommended. In this case the user should select the AAL to use based on its data generation patterns.

6. Future of AAL2

This paper has introduced the fundamentals of AAL2. A bit of thinking shows that AAL2 is self-contained and, therefore, a switched AAL2 network can be defined on the top of ATM as an overlay network. What would be the use for that? Well, with proper signalling AAL2 could be carried over other technologies than ATM, which would allow for a global network hiding the complexity of the lower layers (merely an academic topic nowadays). A more realistic application for AAL2 switching has been identified for the third generation of cellular networks, in the frame of IMT-2000 (International Mobile Telecommunications) standardization, to support a functionality known as soft-handover. But this is a subject for a whole new paper.

Charging of the ABR Service in ATM Networks: a Numerical Example

Llorenç Cerdà, Olga Casals
Computer Architecture Department
Polytechnic University of Catalonia

1. Introduction

The Available Bit Rate service (ABR) is a "best effort" service intended for traffic which imposes no bound on delay or delay variation. The network guarantees a Minimum Cell Rate (MCR) for an ABR source, which is negotiated at the connection set up, and commits to fairly divide the unused bandwidth among all ABR sources.

A usage-based pricing is desirable for an ATM network, the question is how the usage-based scheme should be selected? For sources with guaranteed bandwidth as CBR and VBR, the allocated bandwidth or the generalized concept of the effective bandwidth, and the duration of the connection is a good characterization of their resources usage. Charging of ABR may be more complex because it is responsibility of the network to fairly divide the network resources among the sources, but the network may not have an a priori knowledge of the users appraisal of such resources.

Figure 1. Network topology.

Songhurst and Kelly [6] propose a unified pricing model for VBR and ABR services which consists of fixing the charging parameters at the connection set up, and computing the connection charge by multiplying these parameters by the duration and number of transmitted cells. For the VBR service the charging parameters are computed based on the effective bandwidth, while for ABR the charging parameters are computed based on the requested MCR. The authors assume that the network divides the free bandwidth proportionally to the MCR. MCR is

---

1 This work was partly supported by the Ministry of Education of Spain under grant TIC95-09-82 and by project ACTS 094 "EXPERT".
therefore the mean for the users to communicate their preferences to the network.

We refer the previous approach as a "static" charging model because the charging parameters are established at the connection set up and do not change afterwards. Another approach consists of dynamically varying the charging parameters in order to adapt the source demand to the network capacity. L. Murphy and J. Murphy first proposed this pricing framework for ATM, see for example [3]. Based on the same ideas, a dynamic charging for the ABR service has been proposed by Courcoubetis et al. [1].

In this paper we first overview the static model of Songhurst and Kelly in section 2, and the dynamic model of Courcoubetis et al. in section 3. In section 4 we discuss the pros and cons of the static and the dynamic model and we propose a new pricing model. A numerical comparison of the three pricing models is performed in section 5 where some concluding remarks are given.

2. A Static Pricing Model Based on the Duration of the Connection

Songhurst and Kelly [6] propose a pricing model where the charge of a connection is given by the expression:

\[
\text{Total Charge of a connection} = a(x) \cdot T + b(x) \cdot V + c(x)
\]  

(1)

where \( T \) is the duration of the connection, \( V \) the volume (number of cells) submitted by the connection, and \( x \) is the tariff choice which includes the service class, the traffic contract parameters and others. \( c(x) \) is a fixed subscription fee. Based on the effective bandwidth, for VBR connections Songhurst and Kelly [6] give an expression for \( a(x) \) and \( b(x) \) which minimizes the expected charge given the expected mean and peak rate of the connection. Therefore, the scheme encourages the user to give an accurate value of the traffic contract parameters.

Songhurst and Kelly [6] propose using the expression 1 also for charging ABR sources. In this case \( a(x) \) is assessed proportional to the MCR, i.e. \( a(x) \propto MCR \), and \( b(x) \) is assumed to be much lower than \( y \) or even zero. Note that if the connection transmits always at a rate \( r \geq MCR \), then MCR \( T \) is the volume of traffic submitted at the guaranteed MCR. Therefore, it can be interpreted that the traffic transmitted above the MCR is charged at a much cheaper price. In order for the sources to select a MCR according to their bandwidth appraisal, the authors assume that the network divides the free bandwidth proportionally to the MCR.

3. A Dynamic Pricing Model Based on the Source Demands

Based on the principle of the social welfare optimization [4], a dynamic charging for the ABR Service has been proposed by Courcoubetis et al. [1]. The method consists of using the forward RM-Cells of the ABR service to convey the source demand to the switches, and the backward RM-cells to convey the prices to the sources. Both demand and prices are adjusted by an iterative algorithm, which in equilibrium satisfies the source demand and maximizes the network revenue (maximum social welfare).

For such algorithm the authors in [1] propose adding two new fields to the RM-Cells: a request bandwidth (RB) field and a price per unit of bandwidth (PB) field. Based on the prices \( w_c \), the sources set the RB field with a demand function \( D_c(w_c) \). This value is used by the switches to compute the price \( a_t \) per unit of bandwidth (per unit of time) charged to the connections \( c \) traversing the link \( l \). The switches post the prices to the sources by increasing the PB field of the RM-Cells by \( a_t \). Therefore, denoting the route of a connection \( c \) by \( R_c = \{s_e, a_t\} \), the switches compute the prices periodically at each charging interval \( n \) as:

\[
R_c = \text{Weighted ERICA}
\]

\[
R_c = \text{Weighted ERICA}
\]

\[
\text{Figure 2. Pricing based on time. Switch type: Weighted ERICA}
\]

\[
\text{Figure 3. Pricing based on per cell tariffs. Switch type: Weighted ERICA}
\]
\[
\alpha_n^k = \begin{cases} 
\frac{\sum D_i \alpha_n^k - C_i}{C_i} \alpha_n^k, & \text{if } \alpha_n^k \neq 0 \\
\max \left\{ k \frac{D_i \alpha_n^k - C_i}{C_i} \right\}, & \text{if } \alpha_n^k = 0
\end{cases}
\]

Where \( C_i \) is the capacity available for ABR traffic traversing link \( i \). Oscillations and convergence speed of the algorithm depend on the selection of parameters \( k \) and \( h \). Finally, source charge is computed by a policing and billing unit located at the network access point as Charge = \( \sum w_n V_n^\alpha \), where \( V_n \) is the number of cells sent by the connection during interval \( n \).

4. A Static Pricing Model with Customizable Per Cell Tariffs

The pros and cons of the charging schemes described in the previous sections can be summarized in the following conclusions:

- The main advantage of the static model is the simplicity. The major drawback is that many data connections are very bursty and may not be able to choose a MCR.
- The dynamic model of section 3 solves the previous item and optimizes the network usage from an economical point of view, but has severe drawbacks as:
  - increases the complexity of sources and switches algorithms,
  - increases the complexity of the accounting system which has to keep track the RM-Cells to compute charges,
  - adds charging parameters which need to be tuned for stability,
  - sources pay according to congestion, this may be unfair for sources connected to more congested areas.

We thus believe that a static pricing model would be appropriated, but it should take into account the bursty nature of many data sources. For such sources a pricing model which charges the transmitted volume rather than the duration of the connections would fit better.

The model we suggest would consist of several prices per cell tariffs \( p_i \) for the ABR service. Users would select a tariff \( p_i \) at the connection set up in order to charge the cells transmitted at the shared bandwidth. We further assume that the network allocation algorithm is to divide the free bandwidth proportionally to the chosen tariff \( p_i \). Of course, some sources may need a guaranteed MCR, which is likely to be charged based on the duration of the connection. Thus, we propose the following charging equation:

Total Charge of a connection =
\[
\gamma MCR T + p_i \max\{V - MCR T, 0\}
\]

Note that if the connection transmits always at a rate \( \gamma \geq MCR \), then MCR \( T \) is the volume of traffic submitted at the guaranteed MCR. Therefore, the first term of the right side of equation (3) charges this traffic at \( \gamma \) [unit of price/cell]. The second part of equation (3) is intended to charge the volume of traffic transmitted above the MCR at the price \( p_i \) chosen by the user. \( p_i < \gamma \). Clearly, in the case of a source rate \( \gamma \geq MCR \), the number of cells given by \( \max\{V - MCR T, 0\} \) are the cells transmitted above the MCR. This is not a drawback, however, because it would penalize users who choose a guaranteed MCR higher than their needs. This pricing scheme would have the advantages of the static scheme described in section 2, but would also be appropriate for bursty sources.

5. Numerical Comparison

In this section we perform a numerical comparison of the three pricing schemes described in this paper, namely, the Static Pricing model based on the Duration of the connection (SPD) described in section 2, the dynamic pricing model (DP) described in section 3, and the Static Pricing model based on per Cell tariffs (SPC) proposed in section 4.

We assume the network topology of figure 1, where a greedy (S1), an ON-OFF (O1), and a background of 5 greedy sources (S2) feed a common switch. In order to select the source parameters we assume that the user criterion is to fix a maximum charge per unit of time \( \alpha \). We also assume that the sources are not able to predict their traffic pattern. Finally, we suppose that the sources S1 and O1 are ready to pay double in order to get a higher bandwidth, thus sources S1 and O1 chose \( \alpha = 2 \) and the sources S2 chose \( \alpha = 1 \).

In the SPD model, the price per unit of time is proportional to the MCR. We thus take without loss of generality \( MCR = \alpha \), and the charging formula will be given by the equation (1) with \( a(x) = \alpha, b(x) = 0, c(x) = 0 \). Remember that the switch is assumed to divide the available bandwidth proportionally to the MCR, therefore, the switch applies a weight \( \alpha \) in this scheme. In the DP scheme, being \( w_n \) the price per transmitted cell posted by the network and \( D \), the

Figure 4. Dynamic pricing. Switch type: ERICA.
source rate, the charge after a time period $T$ becomes: $\text{charge} = w \cdot D \cdot T$. Therefore, to upper bound the charge per unit of time by $\alpha$ (i.e. charge$/T \leq \alpha)$, the source demand has to be $D \leq \alpha/w$. The source is upper bounded by the PCR, thus such a source demand function is given by:

$$D_s(w) = \min \left( \frac{\alpha}{w}, \text{PCR} \right)$$

(4)

Finally, in the SPC scheme we assume that the user selects MCR = 0 and upper bound the charge per time unit by choosing a price per cell (and thus a switch weight) equal to $p_s = \alpha/\text{PCR}$ (cfr. equation (3)).

Table 1 summarizes the source parameters corresponding to each pricing scheme. In the simulation we assume that the greedy sources become active at time $t=0$ and remain active until the end of the simulation at $t=1000$ ms. The ON-OFF source becomes active at $t=100$ ms, transmits three bursts of 3000 Cells with a greedy behavior spaced 300 ms and becomes silent after the end of the third burst.

Figures 2, 3 and 4 show the rate evolution of sources O1, S1 and one of the five S2 for the three pricing schemes. In the simulations carried out in figures 2 and 3 we use a weighted switch [5] applying the weights shown in table 1, while in figure 4 the ERICA switch [2] is used performing a max-min fairness. We have to use the ERICA switch because with the weighted switch, the switch and the charging algorithms interfere.

Table 2 shows the measures taken during the simulation applying the three pricing schemes. We define the usage as the ratio between the time while the source is active to the duration of the connection. We also define the expected charge per time as the value $\alpha$ chosen by the source multiplied by the usage. In our framework, this value could be taken as the goal of the charging scheme.

In the case of SPD, the user exactly pays the selected charge per time, regardless the active or idle state of the source. This severely penalizes the bursty source O1, widely exceeding the expected charge per time. In this sense, the DP scheme yields a more fair charge, and all the O1, S1 and S2 sources nearly achieve the expected charge per unit of time. Finally, in the SPC scheme, the charge per cell is fixed. Remember that this has been chosen to be $p_s = \alpha/\text{PCR}$, thus the difference between the measured and expected charge per time will be as high as the difference between the PCR and the average rate. However, this pricing scheme does not penalize the bursty source O1.

References


http://www.cec.dcu.ie/~murphy/publ/publ.html

