FAIR ATM CHARGING WITH CONSIDERATION OF TRAFFIC CHARACTERISTICS AND QOS PARAMETERS

Georg Carle, Thomas Schmidt, and Jochen Seitz

Abstract

Today, public network providers perform usage-based charging of ATM services based on transmitted cells, allocated bandwidth, and connection time. Nevertheless, it is important being able to simplify the process of charging by considering the requested service category and the implied Quality of Service (QoS) that might be negotiated additionally. We present a charging framework and its application on representative services which is based on the following aspects.

First, we differentiate five ATM service categories: Constant Bit Rate (CBR), Real-Time Variable Bit Rate (rt-VBR), Non-Real-Time Variable Bit Rate (nrt-VRB), Unspecified Bit Rate (UBR), and Available Bit Rate (ABR). These categories are refined by traffic parameters. On this basis, we define the total cost of an ATM-connection by the sum of

- the cost for the establishment of the connection;
- the cost for reservation of the connection (depending on the the chosen service class and depending on the traffic descriptors and QoS-parameters;
- and the usage-based costs.

In order to allow fair charging in cases with non-negligible error rates, we propose to allow a distinction of usage-based costs between actual goodput, redundant data used for forward error correction (FEC), and retransmitted data.

Introduction

As widely accepted, ATM is one of the most important networking technologies in private and public networks. While accounting management is not yet a major topic in private networks, it is essential for public service providers. Up to now, many service providers perform charging based on transmitted cells, allocated bandwidth, and connection time. This type of charging is usually used in combination with strict Usage Parameter Control (UPC), where cells that violate the traffic contract are discarded by the service provider. Typically, such services provide strong guarantees for the service quality and in particular very low cell loss rates. However, there are a number of cases in which non-negligible cell loss rates occur, as, e.g., for UBR (Unspecified Bit Rate) services. Conventional charging concepts similar to charging of ISDN services are not appropriate in these cases, and sophisticated support for accounting management and charging will become more and more important as low-cost low-quality ATM services are to be sold. In addition, group communication services impose even more complex requirements for the management and charging. For the establishment of alternative charging concepts, the ability to perform accounting and charging based on the actual amount of transferred information is of high importance.

Requirements for charging

A satisfying charging model has to deal with the following three main aspects of charging:

1. *Economic Aspects*: The model has to differentiate between several types of service and – within these types – between different levels of service quality. Thus, the customer is animated to choose the cheapest type of service to satisfy his needs and the costs can be assigned to the customer, who has produced them.

Georg Carle, Thomas Schmidt, and Jochen Seitz are with the Institute of Telematics, University of Karlsruhe, D-76128 Karlsruhe, Germany

2. *Technical Aspects*: In order to minimise the effort to compute the different tariffs the model has to be as simple as possible. Thus after having distinguished between different types of services there can be a charging model optimised for each type. As cell-based charging in ATM networks is computationally demanding, dedicated hardware-support may be required.
3. *Customer Aspects*: The customer has to understand the bill and must be aware of the costs he produces whenever he is using a service. Further the costs must not differ significantly from the ones the customer is used to, that is the actual telephone tariffing system. This introduces a threshold of acceptance. Costs should be based on goodput. Errors within the network which require additional traffic for error recovery should not lead to increased costs for the customer.

To summarise these three aspects, charging of ATM services should be kept sufficiently simple to be feasible and intelligible, while allowing to differentiate between different types of service.

Related work

Within the Internet, accounting of successfully transmitted user data (goodput) is also important in order to achieve fair charging. General concepts for accounting within the Internet were presented in [MiHR91].

In [MaVa95], a model for congestible resources as found in networks was presented. This model allows to take into account that network resources can be used by more than one user. Increasing the usage of network resources degrades the service quality for all users which are sharing these resources. The model in [MaVa95] allows to perform usage-based billing consisting of direct cost (cost for resource usage without consideration of congestion), and external cost (congestion cost, reflecting additional delay and cost for other users). In [EdMV95], a system (called the billing gateway BGW) was proposed which is capable of charging users for their TCP traffic. While usage accounting within the Internet is based on IP packets, usage accounting in ATM networks should be based on ATM cells and is therefore computationally more demanding. Our approach to hardware-supported management is intended to provide the processing capability needed in ATM networks.

Errors in ATM networks

There are a number of cases in which cell error and loss rates may be sufficiently high in order to justify the additional effort of accounting user goodput. An important case for ATM services which show a significant error and cell loss rate are networks with wireless links. Another potential source for significant errors and losses are low-cost ATM hardware which does not allow to perform accurate traffic control. In addition, significant cell loss rates can occur in cases where a high statistical multiplexing gain is desired for highly bursty sources over high-speed, wide-area connections. In such a scenario, the following problems are responsible for the deficiencies of conservative traffic control algorithms with respect to achieving a high resource utilisation. During transient periods, sources are not able to obtain an accurate view of the current load state. Their view of the network load is always outdated due to the propagation delay. It can not be expected that the actions performed by a source without accurate view of the current load allow to optimise the quality of service parameters (throughput, delay, and loss). Aggressive control algorithms, as for example proposed in [KiFa95], are designed to achieve a high multiplexing gain, at the cost of potentially high losses. While conservative control algorithms may lead to long transient periods, aggressive control algorithms allow to shorten the transient period under certain conditions. Then they reach an equilibrium state faster, where the parameters can be adjusted according to optimality criteria [HeRo95]. One example for a control algorithm that allows to adjust its aggressiveness is the concept of loss-load curves [WiCh91], which can be used to derive strategies for sources in order to maximise throughput or minimise end-to-end delay. In these cases, resource utilisation, throughput and delay can be increased by accepting higher losses.

For multicast ABR services, different service models are possible, which allow similar trade-offs. A conservative multicast service would limit the transmission rate according to the most congested link of a multicast tree, allowing to achieve very low cell loss rates. In contrast, an aggressive ABR multicast service may use higher transmission rates, while producing cell losses at some congested links.

Error control for reliable services

When applications require a higher reliability than offered by the network, protocols with error control mechanisms must be used. For reliable data communication, protocols with Automatic Repeat Request (ARQ) mechanisms are widely used. ARQ mechanisms have several drawbacks when used in high speed wide-area networks, such as high delay in case of errors, and bad scalability for multipoint connections. An alternative approach is Forward Error Control (FEC), which has a number of advantages such as reducing delay and improving scalability for multipoint connections when used in high-speed wide-area networks. Appropriately dimensioned FEC allows to achieve an overall gain, as shown, e.g., by Biersack [Bier93]. However, FEC is not yet widely used over high speed networks, as there still remain a number of important questions such as how to dimension the amount of redundancy and the size of protocol data units. There exist a number of protocols with ARQ or FEC mechanisms which are suitable for ATM networks, and which all have individual strengths and weaknesses.

For all service categories (i.e., CBR, VBR, ABR, UBR), an AAL-level FEC scheme may improve the end-to-end service quality [CaDE95a]. In addition, the implementation complexity in order to implement an AAL-level FEC scheme can be kept relatively small [TsKE95].

An AAL-level FEC scheme allows to provide a packet error rate which is sufficiently small to meet the requirements of the layer above the AAL. Therefore, it enables the efficient provision of scalable reliable multicast over an ATM network, as it may ensure a packet error rate which is sufficiently small to achieve the required end-to-end service quality of multicast scenarios [EsCD95].

For AAL1, an SSCS with FEC is proposed [I363], based on a Reed-Solomon-Code applied on blocks of 128 cells that allows for the regeneration of up to four missing cells. Additional FEC schemes for ATM were proposed and investigated (e.g., [McAu90, ShMc90, OhKi91]), but there are still a number of open questions concerning the combination of FEC and ARQ in ATM networks.

Within the ATM Forum, a specification for a Service-Specific Convergence Sublayer with FEC (FEC-SSCS)was presented [CaEG95a, CaEG95b] which supports cell-based FEC for AAL5. This protocol can be used in combination with all service classes. It allows to adjust the amount of redundancy dynamically.

While FEC allows to correct a certain number of losses originating from congestion, it may also lead to additional cell losses due to the increased load. Therefore, applying FEC is only useful in cases where the loss probability after decoding is lower than the loss rate of a data stream without FEC. As shown in [ShMc90], [Bier93], an appropriately dimensioned amount of redundancy allows to reduce the loss probability after decoding. The Service Specific Connection Oriented Protocol (SSCOP) ([Q2110], [Hen95]) offers an assured mode service using an AAL-level ARQ scheme with selective retransmissions. [EsCa95] discusses improved delay characteristics of SSCOP by the deployment of FEC-SSCS between SSCOP and AAL5-CPCS.

As SSCOP does not support assured mode multicast connections, other protocols are required for reliable multicast connections. The adaptation layer protocol RMC-AAL (Reliable Multicast ATM Adaptation Layer) provides frame-based ARQ, cell-based ARQ and FEC mechanisms for reliable point-to-point and point-to-multipoint services [Carl95, CaSc95, CaZi95]. It is an extension of AAL5 and is suitable for a hardware-based implementation. It can be integrated into end systems and into AAL-level servers. We selected RMC-AAL as platform for an implementation of the fair charging scheme, as AAL-level servers can perform goodput accounting within the network, and are suitable for charging of multipoint services (see Figure 1).

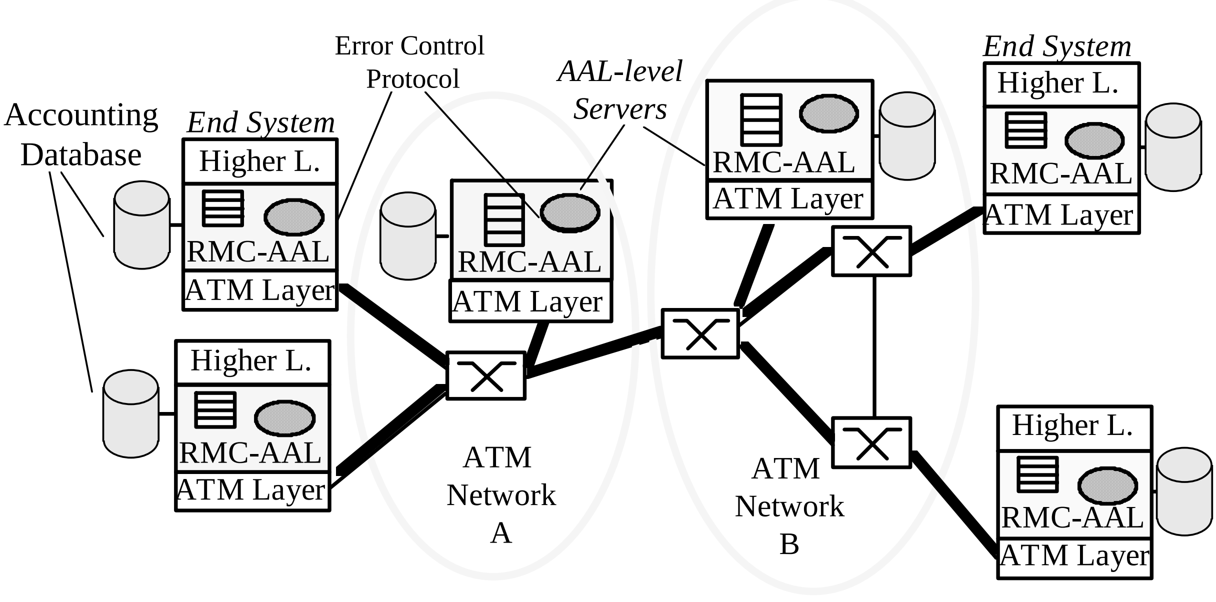

Figure 1: Accounting of user goodput

Influence of Error Control on Charging

For charging, current networks do not distinguish between initially transmitted and retransmitted data. Therefore, usage-based charging may lead to unfairness in the case of a public ATM service with non-negligible cell loss rate, as users have to pay also for retransmitted data. Up to now, neither the network provider nor the user is able to distinguish between initially transmitted data and retransmitted data at the user network interface. Therefore, users currently have no alternative than paying for retransmissions caused by network errors.

The conventional approach for charging is the reservation on peak rate allocation or on sustainable rate in combination with peak rate and burst tolerance. In combination with Call Admission Control (CAC), this enables the service provider to guarantee a certain quality of service. However, in case of highly bursty traffic, this approach allows only a low multiplexing gain and, therefore, leads to relatively high costs.

Our new approach is intended for best-effort services that allow to achieve a high multiplexing gain even for highly bursty traffic, providing a high-performance service at relatively low costs. Our approach aims at charging based on user goodput. This can be realised by counting of cells per connection that are transmitted the first time and distinguishing them from retransmitted cells. Alternatively, user goodput can be evaluated more accurately based on the number of user bytes by exploitation of control information of the error control protocol.

The Proposed Charging Model

For an error control protocol with FEC and ARQ which is known by the network provider, the service provider can use accounting servers which are able to distinguish between the first transmission of a frame, retransmissions of this frame, and redundant data used for FEC. This allows to perform charging with the following, very general charging function:

Cost of ATM connection =
Cost_connection (service class, traffic parameters, distance)
\+ **Cost_reservation** (service class, traffic parameters, distance) · #cell_times
\+ **Cost_ goodput** (service class, traffic parameters, distance) · #cells
\+ **Cost_ retransm.** (service class, traffic parameters, distance) · #cells
\+ **Cost_ redundancy** (service class, traffic parameters, distance) · #cells

Figure 2: Fair charging function for services with non-negligible loss rates

In Figure 2, one of the five ATM service classes can be used: Constant Bit Rate (CBR), Real-Time Variable Bit Rate (rt-VBR), Non-Real-Time Variable Bit Rate (nrt-VRB), Unspecified Bit Rate (UBR), and Available Bit Rate (ABR). These categories are refined by traffic parameters, consisting of traffic descriptors [Peak Cell Rate (PCR), Cell Delay Variation Tolerance (CDVT), Sustainable Cell Rate (SCR), and Burst Tolerance (BT)] and QoS parameters [Cell Loss Ratio (CLR), Mean Cell Transfer Delay (Mean CTD), Maximum Cell Transfer Delay (Max CTD), and Cell Delay Variation (CDV)].

The cost for reservation in Figure 2 consists of the cost for the forward direction of the connection (calling to called party) in terms of number of switches, amount of data that has been transmitted, and a tariff for the chosen service category depending on the traffic parameters, and the corresponding cost for the backward direction (called to calling party). It is proportional to the duration time of the connection.

Usage-based costs consist of the cost for actual goodput, cost for retransmissions, and cost for redundancy. Cost for goodput represent the major part of usage costs. Cost for retransmission may be set to zero, or to an amount which prevents fraud, and which also motivates users to minimise errors. Cost for redundancy will be higher than cost for retransmissions, but lower than cost for goodput.

This charging function allows to accurately taking into account user goodput by distinguishing between user data, retransmissions, and redundancy. Such a charging scheme has the disadvantage of being fairly complex, but has the following advantages. It encourages users to deploy processing capability required for FEC. It also gives users a stronger correlation between the goodput they achieve, and the cost associated with this service.

If usage-based costs are to be separated between goodput, retransmissions and redundancy in scenarios with non-negligible errors, then the protocol deployed by the users has to be known to the service provider, and has to be processed by dedicated accounting servers. In case a user deploys a protocol for error recovery which is not known to the service provider, all his traffic will be treated as goodput.

Implementation of accounting management

An implementation of the proposed charging scheme has to meet the following requirements. One is the easy access to collected data from management software, the other is the capability for high-performance processing needed for counting on cell or frame level. The appropriate component can be located at the customers network interface (customer or provider site) or, alternatively, within the public network for performance optimisation.

For management purposes we started to define an appropriate SNMP management information base (MIB). The following small example from this MIB describes a counter for cells with user data which are transmitted the first time, which allows to measure user goodput :

```
fairATMChargingCellCounter OBJECT-TYPE
SYNTAX Integer64
MAX-ACCESS read-only
STATUS current
DESCRIPTION "Cells with user data which are transmitted for the first time"
::= { fairATMChargingCounters 1}
```

High-performance Processing of Accounting Management

Advances in protocol implementation [Stee94, KrKS93] and hardware support for protocol processing [CaSc95b, Schi95] are beginning to overcome the performance bottleneck in end systems, supporting reliable high-performance services over ATM networks. However, there is still a lack of support for efficient management and charging of high performance communication services.

The implementation of the hardware-component that collects the appropriate values for the MIB is done using the modular hardware-architecture CHIMPSY, which was especially designed for high-performance communication processing [Schi95]. Our concept for hardware-supported accounting management uses a dedicated processing component, the Management Processing Unit (MPU). This component will be an additional element of our hardware library, which provides RISC-kernels, timers, and other dedicated components for communication. Using these components, we developed a hardware architecture called the Generic ATM Protocol Processing Unit (GAPPU) [CaSc95]. Attached to the central crossbar of the GAPPU, the management processing unit listens to the cell stream and updates data structures containing per connection cell counters. A new counter is initialised at connection setup. The component can now perform operations like increment, read, and delete the counter. It provides SNMP-like primitive services to the management proxy agent which has full SNMP functionality and is executed by a host CPU. Communication with the management proxy agent is performed using the following MPU functions:

```
mpu_set (connection_id, parameter_id, value)
mpu_get (connection_id, parameter_id)
mpu_trap (connection_id, parameter_id, value)
```

This architecture provides ATM end and intermediate systems with functionality and processing power as needed for accurate and fair cell-based charging. In particular, the architecture allows fair charging for ATM services with non-negligible cell losses.

Conclusions

Main benefits of our approach are the possibilities for users to control traffic, and for the providers to perform accurate and fair charging. Furthermore, network providers can easily distinguish between retransmissions caused by errors inside the public network and errors caused by end-systems (buffer overflow etc.), by simply comparing the incoming cell counts of the network interface at the sender side with the outgoing cell count at the receiver side. This first step towards fair charging allows the provision of high-performance best-effort services which are more cost-effective than traditional, conservative services that provide quality of service guarantees.

The presented architecture for accounting management allows to trade off network costs (i.e., costs for consumed resources of the network), which are related to the total number of transmitted cells, and user benefit (i.e., costs considered fair by the user), which are related to goodput (in terms of received cells, or number of user bytes). The architecture allows to monitor these values, and to store them in managed objects of an accounting MIB. By supporting fair charging, the architecture motivates users to provide protocol processing capability in the end systems which are required to recover losses which originate from a high degree of resource sharing within the network.

References

[Bier93] Biersack, E.: *Performance Evaluation of Forward Error Correction in an ATM Environment,* IEEE Journal on Selected Areas in Communication, Vol. 11, Nr. 4, May 1993, pp. 631-640

[CaEG95a] Carle, G.; Esaki, H.; Guha, A.; Tsunoda, K.; Kanai, K.: *Necessity of an FEC Scheme for ATM Networks,* Contribution ATMF/95-0325; ATM Forum Technical Committee ’Service Aspects and Applications SA&A’, Denver,U.S.A., April 1995

[CaEG95b] Carle, G.; Esaki, H.; Guha, A.; Tsunoda, K.; Kanai, K.: *Proposal for Specification of FEC-SSCS for AAL Type 5,* Contribution ATMF/95-0326; ATM Forum Technical Committee ’Service Aspects and Applications SA&A’, Denver,U.S.A., April 1995

[Carl95] Carle, G.: Towards Scalable Error Control for Reliable Multicast Services in ATM Networks, Twelfth International Conference on Computer Communication, ICCC’95, Seoul, Korea, August 20-25, 1995

[CaSc95] Carle, G.; Schiller, J.: *Enabling High Bandwidth Applications by High-Performance Multicast Transfer Protocol Processing,* 6th IFIP Conference on Performance of Computer Networks, PCN95, Istanbul, Turkey, October 1995

[CaZi95] Carle, G.; Zitterbart, M.: *ATM Adaptation Layer and Group Communication Servers for High-Performance Multipoint Services,* 7th IEEE Workshop on Local and Metropolitan Area Networks, Marathon, Florida, March 1995

[EdMV95] Edell, R.; McKeown, N.; Varaiya, P.: *Billing Users and Pricing for TCP,* IEEE Journal on Selected Areas in Communications, Volume 13, Number 7, September 1995, pp. 1162-1162

[EsCa95] Esaki, H.; Carle, G.: Combination of SSCOP and an AAL-Level FEC Scheme, ATM Forum Document ATMF/95-1560, ATM Forum Plenary, London, U.K., 11.-15. Dezember 1995

[EsCD95] Esaki, H.; Carle, G.; Dwight, T.: *Reliable Multicast Service Needs Cell-Level FEC Scheme,* ATM Forum Document ATMF/95-1438; ATM Forum Plenary, London, U.K., Dezember 1995

[HeRo95] Heinrichs, B.; Rossmann, M.*: ABR: The ATM Service for Data Communication Applications?,* First Workshop on ATM Traffic Management (WATM‘95), Paris, France, December 1995

[I363] ITU-TS Draft Recommendation I.363: BISDN ATM Adaptation Layer (AAL) Specification, Geneva, 1993

[KiFa95] Kim, H.; Farber, D.: *New Congestion Control Framework for Large Bandwidth-delay Product Networks,* Proceedings of 6th International Conference on High Performance Networking (HPN'95), Palma de Mallorca, Spain, September 1995

[KrKS93] Krishnakumar, A.S.; Kneuer, J.G.; Shaw, A.J.: *HIPOD: An Architecture for High-Speed Protocol Implementations,* in: Danthine, A.; Spaniol, O. (Eds.): High Performance Networking, IV, IFIP, North-Holland, 1993, pp. 383-396

[MaVa95] MacKie-Mason, J.; Varian, H.: *Pricing Congestibel Network Resorces,* IEEE Journal on Selected Areas in Communications, Volume 13, Number 7, September 1995, pp. 1141-1149

[McAu90] McAuley, A.: Reliable Broadband Communication Using a Burst Erasure Correcting Code, ACM SIGCOMM '90, Philadelphia, PA, U.S.A., September 1990

[MiHR91] Mills, C.; Hirsch, D.; Ruth, G.: *Internet Accounting: Background,* Internet Engineering Task Force, RFC 1272, November 1991

[OhKi91] Ohta, H.; Kitami, T.: A Cell Loss Recovery Method Using FEC in ATM Networks, IEEE Journal on Selected Areas in Communications, Vol.9, No.9, Dec. 1991, pp.1471-1483

[Schi95] Schiller, J.: *A Flexible Co-Processor for High-Performance Communication Support,* IEEE Globecom'95, Singapore, November 1995

[ShMc90] Shacham, N.; McKenny, P.: Packet recovery in high-speed networks using coding, Proceedings of IEEE INFOCOM '90, San Francisco, CA, pp. 124-131, June 1990

[Stee94] Steenkiste, P.: *A Systematic Approach to Host Interface Design for High-Speed Networks,* IEEE Computer, March 1994, pp. 47-57

[TsKE95] Tsunoda, K.; Kanai, K.; Esaki, H.: *A Reed-Solomon Erasure Code and Its Application to AAL,* ATM Forum Document 95-1162, ATM Forum Plenary, Toronto, Canada, August 1995

[WiCh91] Williamson, C.; Cheriton, D.: *Loss-Load Curves: Support for Rate-Based Congestion Control in High-Speed Datagram Networks,* Proceedings of ACM SIGCOMM 1991, Zurich, Switzerland, September 1991, pp. 17-28