TRAFFIC MANAGEMENT IN ATM NETWORKS : AN OVERVIEW [1]

C. Blondia† and O. Casals‡

† *University of Antwerp*
Department of Computer Science and Mathematics
Universiteitsplein 1 , B-2610 Antwerpen, Belgium
blondia@uia.ua.ac.be

‡ *Polytechnic University of Catalunya*
Computer Architecture Department
c/ Gran Capitan, Mod. D6, E-08071 Barcelona, Spain
olga@ac.upc.es

Abstract

The main objectives of traffic management in ATM networks are to protect the user and the network in order to achieve network performance objectives and to use the available resources in an efficient way. In order to achieve these objectives the profile of the cell stream of each connection needs to be described adequately by means of a set of traffic parameters, together with an indication of the required level of QoS. The relationship between network performance and traffic characteristics and QoS is structured by means of ATM layer Service Categories and Transfer Capabilities. Each Category/Capability is provided with a number of traffic congestion and traffic control mechanisms needed to guarantee the required QoS of the category while achieving a high level of efficiency. This paper presents a state-of-the-art of traffic management in ATM networks. An overview is given of the Service Categories, together with the most important control and congestion schemes : CAC, UPC, traffic shaping, priority control, resource management, flow control, packet discarding schemes.

1. Introduction

The Asynchronous Transfer Mode (ATM) has been chosen as the transfer mode for B-ISDN because of its flexibility to support various types of services, each having their own traffic characteristics and performance requirements, and because of its efficiency with respect to resource utilisation, due to the potential gain by statistically multiplexing bursty traffic. Since ATM has to provide differentiated Quality of Service (QoS) to the various applications, there is a need for efficient, effective and

[1] This work was supported by the European Union under project AC094 (EXPERT). The first author was also supported by Vlaams Actieprogramma Informatietechnologie under project ITA/950214/INTEC (Design and control of broadband networks for multimedia applications). The second author was supported by the Spanish Ministry of Education under project TIC95-0982-C02-01.

simple functions which control the traffic streams and their resource utilization. These ATM layer traffic and congestion control functions are referred to as *Traffic Management* mechanisms. They are defined and standardised by ITU-T in Recommendation I.371 (Traffic Control and Congestion Control in B-ISDN, see [I371]) and by the ATM Forum in Traffic Management Specification 4.0 (see [ATM95]). The objective of traffic management is twofold.

- To achieve well-defined *performance objectives* by protecting both the user and the network against congestion. These performance objectives can be expressed in terms of cell loss probabilities, cell transfer delay, cell delay variations, etc.
- To achieve *efficiency* and *optimisation* of the usage of network resources needed to ensure the above mentioned performance requirements.

Traffic management mechanisms should be able to take the appropriate actions under all possible traffic conditions, such as

- temporarily *overload conditions* due to the statistical fluctuation of variable bit rate traffic
- *malicious users*, who deliberately offer more traffic to the network to obtain operational and/or economical advantage with respect to the other users
- *malfunctioning* of terminal equipment, leading to unexpected traffic volumes entering the network.

In order to structure the relationship between traffic characteristics and QoS requirements on one hand and network behaviour on the other hand, ATM Service Categories (ATM Forum terminology) or ATM Transfer Capabilities (ITU-T terminology) have been introduced. These service categories are intended to support a number of ATM Service Classes and associated QoS by means of a set of appropriate traffic management mechanisms.

The aim of this paper is to give an overview of these mechanisms. It is structured as follows. In Section 2 the parameters needed to define the notion of QoS and to characterize the traffic are introduced. Section 3 gives an overview of the ATM Service Categories and ATM Transfer Capabilities currently defined or under definition. In Section 4, we discuss the most important traffic control mechanisms : CAC, UPC/NPC, traffic shaping, priority control and resource management mechanism. Section 5 deals with congestion control mechanisms for Best Effort type of service. Here we discuss the ABR flow control scheme, several intelligent packet discarding schemes for the UBR Service Category and the mechanisms related to the Guaranteed Frame Rate Service Category. Finally conclusions are drawn in Section 6.

2. Parameters of Quality of Service and Traffic Characterization

2.1 ATM Layer Quality of Service

The ATM layer Quality of Service (QoS) is defined by means of a set of parameters which characterize the end-to-end performance of a connection at the ATM layer. These parameters can be divided into two classes, namely parameters which may be negotiated between the end-systems and the network and parameters which are given by the network. There are three parameters which may be negotiated, two related to cell delay and one related to cell loss. In order to better understand these definitions, we use the cell transfer delay probability density function depicted in Figure 1.

The *Maximum Cell Transfer Delay* (maxCTD) is defined to be the (1-α) quantile of the Cell Transfer Delay (CTD). The *Peak-to-peak Cell Delay Variation* (Peak-to-peak CDV) is defined to be the (1-α) quantile of the CTD minus the fixed CTD (which represents the component of the delay due to propagation and switch processing). This measure quantifies the difference between the best case of CTD (i.e. the fixed CTD) and the possible worst case of CTD (i.e. the CTD value which is not exceeded with a probability smaller than α). The *Cell Loss Ratio* (CLR) is defined to be the number of lost cells divided by the total number of transmitted cells. Remark that lost cells include those that are delivered late w.r.t. the (1-α) quantile of the Cell Transfer Delay.
There are three non-negotiated QoS parameters. The *Cell Error Ratio* (CER) is defined as the ratio between the number of errored cells and the total number of successfully transferred cells and errored cells. The *Severely-Errored Cell Block Ratio* (SEBR) is the number of severly errored cell blocks (i.e. when more than M cells are errored, lost or misinserted in a block of N cells) divided by the total number of transmitted cell blocks. Finally, the *Cell Misinsertion Rate* (CMR) is defined to be the ratio between the number of misinserted cells during a time interval and this time interval.

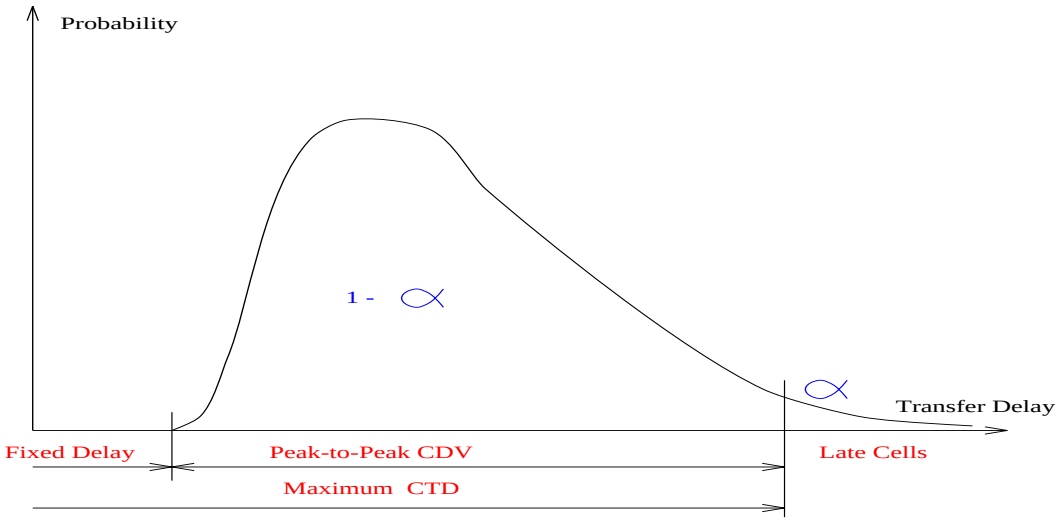


Figure 1 : Cell Transfer Delay Probability Density

2.2 Traffic Parameters and the Generic Cell Rate Algorithm

Traffic parameters are used to describe traffic characteristics of an ATM connection. These characteristics may be qualitative (e.g. telephony, data transfer) or quantitative (e.g. the peak cell rate value). The traffic parameters are grouped in an *ATM traffic descriptor*. The subset of the traffic descriptor that is used during the connection establishment is called the *source traffic descriptor*. A major requirement of an ATM traffic parameter is its suitability to test whether a connection behaves conform the values of this parameter. Therefore, these parameters are given an operational definition, rather than a statistical definition. This means that they are defined according to an algorithm which allows conformance testing in a direct way, opposite

to e.g. the mean bit rate, a statistical definition. The algorithm used to define the traffic parameters in an operational way is the *Generic Cell Rate Algorithm* (GCRA). There are two equivalent definitions of the GCRA, namely the *Virtual Scheduling Algorithm* and the *Continuous Leaky Bucket Algorithm*. We give both definitions and leave it to the reader to check the equivalence. GCRA is defined by means of two parameters T and τ, T being the *increment* and τ being the *limit* and is denoted by GCRA(T, τ).

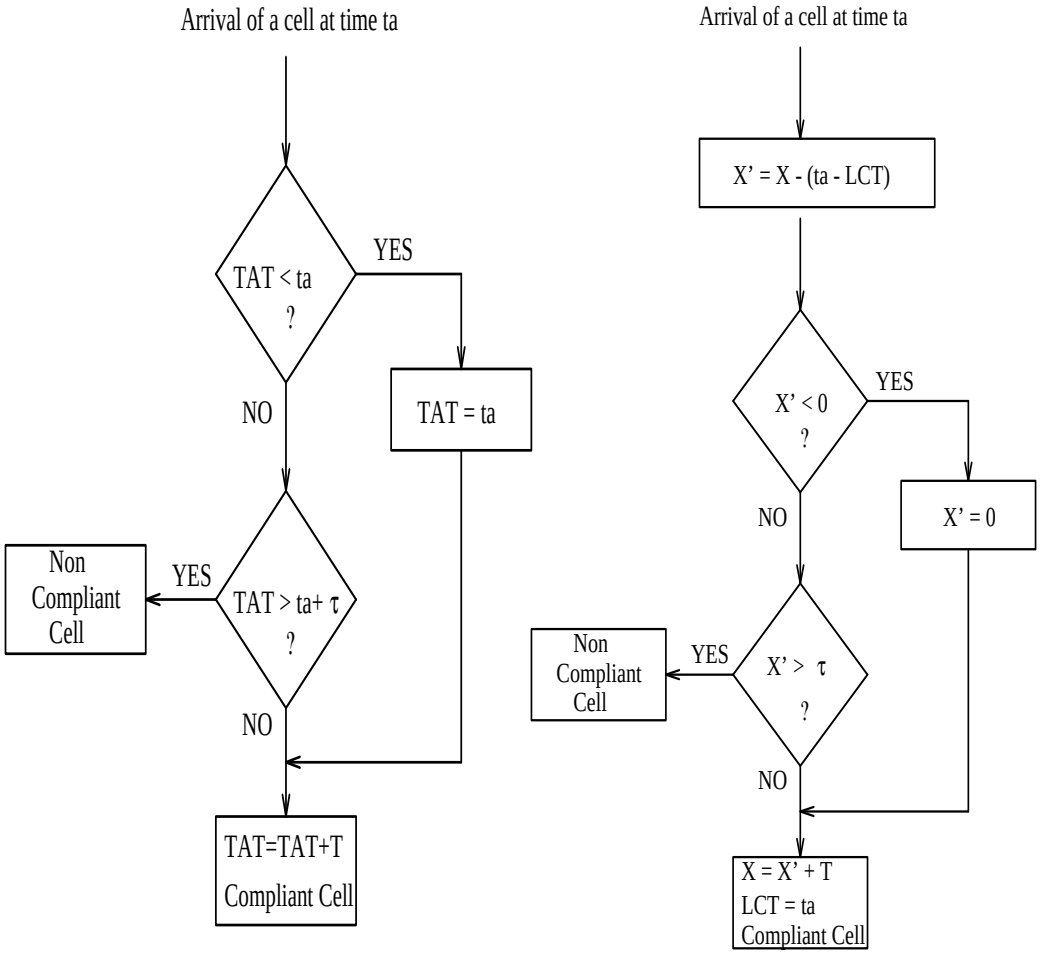


Figure 2 : Generic Cell Rate Algorithm Definitions

The Virtual Scheduling Algorithm (VS)

Let a cell arrive at time t_a. The algorithm computes the Theoretical Arrival Time (TAT) based on the assumption that cells arrive equally spaced (the interarrival time being T). If the previous TAT is smaller than the actual arrival time (meaning that the cell arrived later than theoretically expected), then the cell is declared compliant and the new TAT is computed as t_a + T. If the actual arrival time is greater than TAT-τ (τ representing a certain tolerance), then the cell is conforming and the new TAT is set to

TAT+T. If on the other hand the arrival time is less that TAT-T, then the cell is considered as non-compliant (meaning that the cell arrived too early). The VS algorithm is shown in Figure 2.

The Continuous-State Leaky Bucket

The continuous-state LB is a finite capacity queue (equal to T+τ) with a continuous leak of 1 and of which the content increases by T every time a cell arrives. Its operation is depicted in Figure 2. X denotes the contents of the LB, while LCT denotes the Last Conformance Time. By using the intermediate variable X' it is possible to test whether at cell arrival time the bucket contents level is less than τ, in which case the cell is considered as compliant.

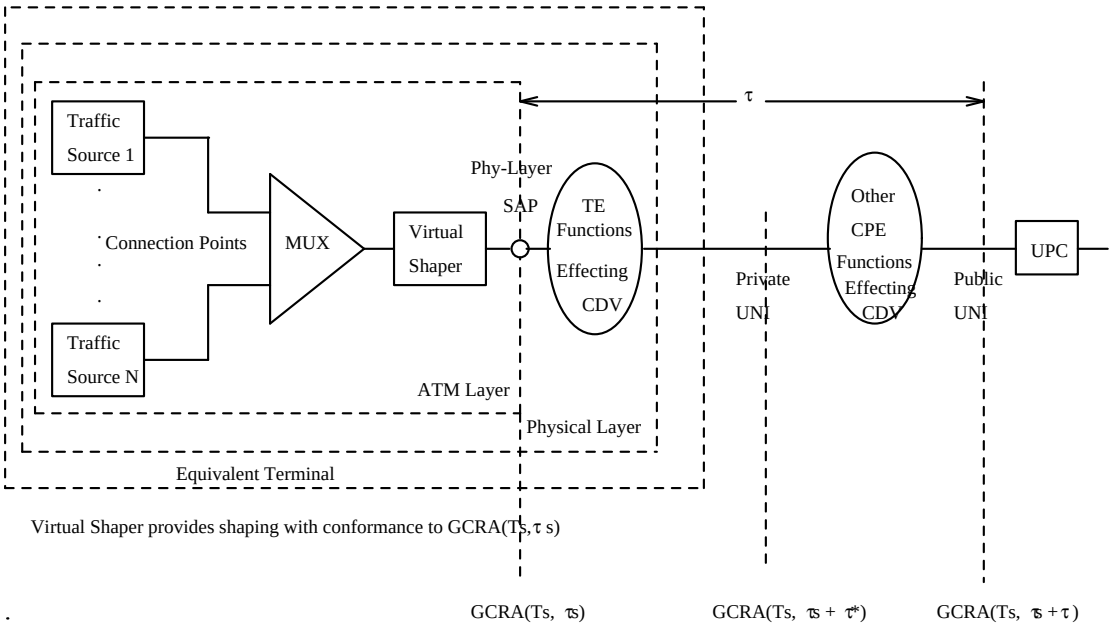


Figure 3 : Reference Configuration

2.3 The Connection Traffic Descriptor

The connection traffic descriptor consists of two parts :

- the source traffic descriptor, being
  - the peak cell rate (PCR)
  - the sustainable cell rate (SCR)
  - the burst tolerance (BT)
- the cell delay variation tolerance (CDVT).

As the cell delay variation tolerance is closely related to the peak cell rate, just as the burst tolerance is related to the sustainable cell rate, we will treat them in another order. For the definition of these parameters we use the reference configuration depicted in Figure 3.

The Peak Cell Rate

The *Peak Cell Rate* (PCR) R_p of a connection is defined at the Physical Layer Service Access Point (SAP), as the inverse of T, the minimum time between emission of two cells from this connection. As different connections may be multiplexed before the Physical Layer SAP, a virtual traffic shaping function is introduced in the reference configuration of Figure 3, to allow a definition of the PCR independent from other connections.

The Cell Delay Variation (CDV) Tolerance

The cell stream of a connection may experience variable delay before entering the network (i.e. before the T_B interface), and hence before being submitted to the policing function. This Cell Delay Variation (CDV) is due to

- ATM Layer functions : when different connections are multiplexed, due to the asynchronous nature, cells of some connections may be delayed. This delay is variable.
- Physical Layer Functions and the insertion of OAM cells
- Customer equipment.

Due to the CDV that is introduced, the UPC function can not operate purely on basis of the PCR. Some tolerance to cope with the CDV has to be built in. In order to define this tolerance, we use the Generic Cell Rate Algorithm.

The CDV tolerance τ, is defined as the second parameter in the Generic Cell Rate Algorithm GCRA(T,τ), where T denotes the inverse of the peak cell rate defined above.

The Sustainable Cell Rate

The PCR (together with the CDV tolerance) describes the cell rate of a CBR connection in an adequate way. However, an important part of the traffic carried by an ATM network consists of VBR traffic (e.g. video). If one would restrict the traffic descriptor to PCR, then resources could only be allocated on basis of the peak cell rate, and no statistical gain could be achieved. Hence a parameter is needed which reflects a kind of average bandwidth utilization of a connection. Since the mean peak rate is not suited for policing purposes (see [RAG]), we define the *Sustainable Cell Rate* (SCR) R_s as the inverse of T_s which takes a value between the minimal cell interarrival time T and the mean cell interarrival time. The sustainable cell rate is defined at the Physical Layer SAP (see Figure 3).

The Burst Tolerance

The Burst Tolerance τ_s is defined as the second parameter in the Generic Cell Rate Algorithm GCRA(T_s, τ_s), where T_s denotes the inverse of the sustainable cell rate defined above. It gives an upper bound on the length of a burst transmitted at peak cell rate. It is easy to show that the maximal burst size B, given T, T_s and τ_s, satisfies $B = 1 + \left\lfloor \frac{\tau_s}{T_s - T} \right\rfloor$, where $\lfloor r \rfloor$ denotes the largest integer value less than or equal to r.

Remark that when a connection has generated a burst at peak cell rate with length B, it has to be idle for a while before generating another burst (the length of the idle period depends on the size of the next burst and the cell rate at which this burst is

generated). Hence, while the PCR and the CDV tolerance control the peak cell rate of a connection, the SCR and the burst tolerance control the burstiness of a connection.

2.4 Compliant Connection

Once the four parameters, PCR, CDV tolerance, SCR and burst tolerance are fixed for a connection, then this connection is said to be compliant whenever the portion of non-compliant cells w.r.t. the total amount of cells generated by the connection, does not exceed a certain value, to be specified in the traffic contract. For compliant connections, the network has to guarantee the agreed QoS, while for non-compliant connections, there is no guarantee at all.
Remark that there are many traffic patterns that are characterized by the same four parameters. In [SKL94] the worst case traffic pattern, satisfying the above four parameters is characterized.

2.5 Traffic Contract

During the connection set-up, a traffic contract between the user and the network is negotiated. This contract contains

- the requested QoS class : these classes are defined using the delay and cell loss parameters defined in 2.1.
- the traffic descriptor : the source traffic descriptor (PCR, SCR, BT) and the CDVT as defined in 2.3.
- the definition of a compliant connection : conformity is defined by means of one or more GCRAs.

3. ATM Service Categories

In order to support efficiently the various services and applications with their specific QoS requirements, a number of ATM Service Categories (ATM Forum terminology) or ATM Transfer Capabilities (ITU-T terminology) have been defined. For each Service Category a set of appropriate traffic control and congestion control functions have to be identified, in order to achieve the required QoS of each class.
The ATM Forum has identified the following classes : Continuous Bit Rate (CBR), real-time Variable Bit Rate (rt-VBR), non-real-time Variable Bit Rate (nrt-VBR), Available Bit Rate (ABR), Unspecified Bit Rate (UBR) and currently under definition Guaranteed Frame Rate (GFR). The ITU-T defines a similar structure, with the exception that no difference is made between real-time and non-real-time VBR, CBR is called Deterministic Bit Rate (DBR), VBR is called Statistical Bit Rate (SBR), UBR and GFR are not defined, but on the other hand the ATM Block Transfer (ABT) Transfer Capability is defined, both with delayed transmission (ABT/DT) and with immediate transmission (ABT/IT). In what follows we give a short description of these categories/capabilities.

3.1 The Continuous Bit Rate Service Category (CBR)

The CBR Service Category, called Deterministic Bit Rate (DBR) by ITU-T, is intended for connections with stringent time relationship and bounded transfer delay and cell transfer delay variation requirements, which need a fixed amount of bandwidth for the whole duration of the connection. This bandwidth is characterized by the Peak Cell Rate. Typical applications using this Service Class are telephony, constant bit rate video and circuit emulation services. The traffic parameters used for this class are PCR and CDVT. The QoS parameters are CLR, peak-to-peak CDV and maxCTD.

3.2 The Real-Time Variable Bit Rate Service Category (rt-VBR)

This Service Category is used for traffic streams with stringent time constraints (as CBR) but which transmit their information at a rate that varies in time. As such, they exhibit a bursty character and hence are suited for statistical multiplexing gain. Typical applications are voice with silence detection and VBR video. The traffic parameters used for this class are PCR, SCR and BT. The QoS guarantees given are CLR, peak-to-peak CDV and maxCTD.

3.3 The non-Real-Time Variable Bit Rate Service Category (nrt-VBR)

This Service Category is meant for non-real-time applications which exhibit a bursty character. As there are no timing constraints, they are very well suited to achieve a high statistical multiplexing gain. Typical applications using this Service Category are response time critical transaction processing such as airline reservations and banking transactions. The traffic parameters that are used are PCR, SCR and BT. The only QoS guarantee is the CLR.

3.4 The Available Bit Rate Service Category (ABR)

The Available Bit Rate Service Category (ABR) has been introduced to support connections originating from users which are willing to accept unreserved bandwidth and which are able to adapt their cell rate to changing network conditions and available resources. Information about the state of the network (e.g. with respect to congestion) and the availability of resources is sent to the source as feedback information through special control cells, called Resource Management cells (RM cells). Services which are compliant to this feedback control information experience a low cell loss ratio and obtain a fair share of the available bandwidth. There is no guarantee with respect to the delay or delay variation. As this control scheme operates at the time scale of a complete round trip delay, the ABR Service Category requires large buffers to be present in the network. The traffic parameters used for this category are the Peak Cell Rate and a minimal usable bandwidth, called the Minimum Cell Rate (MCR). The only QoS guarantee is the CLR. The available bandwidth may vary in time, but shall never be lower than the MCR. Typical applications using this

category are Remote Procedure Calls, Distributed File Transfer, Computer Process Swapping, etc.

3.5 The Unspecified Bit Rate Service Category (UBR)

The Unspecified Bit Rate Service Category (UBR) is meant for traditional computer communication applications (such as e-mail, file transfer, etc.), where no specific QoS guarantees are required. It is the Best Effort ATM Service Category. No guarantees are offered with respect to CLR or CTD. The source will specify a PCR.

3.6 The ATM Block Transfer Capability (ABT)

The ATM Block Transfer Capability (ABT), defined by ITU-T but not considered by the ATM Forum, provides a service with transfer characteristics negotiated on an ATM block basis. As such, it can be considered as a “non-permanent” CBR service. When a block is accepted by the network, sufficient network resources are allocated such that the QoS guarantees are equivalent to those offered to a CBR connection with the PCR negotiated for the transmission of a block. There are two variants of the ABT Transfer Capability : with Delayed Transmission (ABT/DT) and with Immediate Transmission (ABT/IT). In the first case an ATM block is transmitted only after the block cell rate has been confirmed by the network (i.e. after the network has reserved the required resources to transmit the block according to the agreed QoS). In the second case the block is transmitted immediately without waiting for the acknowledgement. This may result in a loss of the whole block if one ore more network elements on the path are short of resources. The traffic parameters specified by the source are PCR, SCR and BT. The QoS guarantees are the CLR, CTD, CDV and the blocking probability.

3.7 The Guaranteed Frame Rate Service Category

The Guaranteed Frame Rate Service Category (GFR) was first proposed in [GUE96] with a different name (UBR+) and is currently under definition by the ATM Forum. The objective of this new service is to incentive users to migrate to ATM technology. The current ATM services require equipment and applications able to describe the traffic parameters needed to establish an ATM connection. However many existing users are not able to specify the required traffic parameters or do not have the suitable equipment. For these users the only possibility to access ATM networks is through UBR connections which do not give any of the ATM QoS guarantees. Therefore the main goal of the GFR service is to bring the benefits of ATM performance and service guarantees to such users. GFR keeps the simplicity of UBR while providing some added features.The GFR service provides the user with a minimum cell rate (MCR) guarantee as long as the user sends frames of size less than the specified value. The service also allows the user a fair share of the spare bandwidth, i.e. the excess traffic of each user will get a fair access to the available resources. The traffic parameters used by GFR are the PCR, CDVT, MCR and the maximum AAL5-PDU size

3.8 Traffic Parameters and QoS Parameters for the Service Categories

The following table summarizes the traffic parameters used and the QoS parameter guaranteed by the different Service Categories introduced in this Section. GFR is not included as it is still under definition.

	CBR	**rt-VBR**	**nrt-VBR**	**ABR**	**ABT**	**UBR**
Traffic Parameter						
PCR,CDVT	specified	specified	specified	specified	specified	specified
SCR, MBS	not applic.	Specified	specified	not applic.	not applic.	not applic.
MCR	not applic.	not applic.	not applic.	specified	not applic.	not applic.
QoS Parameter						
CLR	specified	specified	specified	specified	specified	unspecified
peak-to-peak CDV	specified	specified	unspecified	unspecified	specified	unspecified
maxCTD	specified	specified	unspecified	unspecified	specified	unspecified
blocking prob.	not applic.	not applic.	not applic.	not applic.	specified	not applic.

4. Traffic Control Mechanisms

The basic ATM control functions we discuss in this section are :

- Connection Admission Control (CAC)
- Usage/Network Parameter Control (UPC/NPC)
- Priority Control and Selective Cell Discarding
- Traffic Shaping
- Resource Management

4.1 Connection Admission Control (CAC)

According to ITU-T Recommendation I.371, *Connection Admission Control* (CAC) is the set of actions taken by the network at the call set-up phase (or during the call re-negotiation phase) in order to establish whether a VC/VP connection can be accepted or rejected. A connection is to be accepted at its required Quality of Service (QoS) while maintaining the agreed QoS of already existing connections (for a discussion of QoS, we refer to Section 2.1). The decision whether a connection is accepted or not depends on the network resources that are available (and hence on the load of the network) and on the characteristics of the connection to be established. Hence the following information has to be available to perform CAC :

- The network resources available for new connections

- The QoS level the new connection requires (see Section 2.2)
- The traffic volume generated by the new connection characterized by a *connection traffic descriptor*.

Connection Admission Control for CBR Traffic

When the traffic that is offered to a multiplexer has a constant bit rate, than a straightforward approach could be simply admit connections as long as the sum of the PCRs does not exceed the capacity of the link. The buffer behavior can then be evaluated using the $N \times D/D/1$ or the $\sum N_i \times D_i/D/1$ model (see [VR89] and [RV91]). However, the presence of CDV makes this simple rule not necessary valid, unless the CDV that is allowed is negligible (see [COST242], Section 5.1.1 for a discussion on negligible CDV). When the CVD is not negligible, one may keep the constraint that $\sum_i PCR_i \le C,$ for a link with capacity C, in addition to the condition that $\sum_i b_i \le B,$ when b_i is the burst size of source i and B is the buffer capacity of the multiplexer. Based on those conditions, when C, B and PCR are given, the bucket depth b_i for source i has to be limited by $b_i = r_i \frac{B}{C}$. Note that worst case assumptions are supposed in this model.

Connection Admission Control for VBR Traffic

Assume that a number of VBR sources are to be multiplexed on a link. When the buffer of the multiplexer is intended to absorb cell scale congestion we refer to this type of multiplexer as *Rate Envelope Multiplexing* (REM). If the buffer capacity is large enough to cope with burst scale congestion, we refer to *Rate Sharing Multiplexing* (RSM). In what follows we discuss these two multiplexing schemes and related CAC algorithms in more detail.

Rate Envelope Multiplexing (REM)

When dealing with services which have to meet strict delay requirements, such as interactive voice and video, small buffers (of the order of 100) able to absorb cell scale congestion (i.e. congestion due to a concentration of cell arrivals from different sources) are sufficient. The aim of CAC in this case is to limit the arrival rate such that the probability that the arrival rate exceeds the service rate is negligible. This type of multiplexing is also called *bufferless multiplexing*. With respect to the multiplexing efficiency, studies have shown that REM is efficient for bursty sources with peak cell rates which are low with respect to the link rate. The key idea is to define a notion of *Effective Bandwidth* which is used by the CAC algorithm. To determine the value of the Effective Bandwidth of a source, one may use statistical knowledge about the source (e.g. mean, variance) or one may assume worst case assumptions based on the traffic parameters defined by one or more GCRAs, or one may even use on line measurements. Examples of Effective Bandwidth definition based on statistical characteristics may be found in [KEL91], where a Chernoff bound is used to compute the probability of resource saturation. In [ROB92], an empirical expression is used to determine the Effective Bandwidth based on the mean and the variance of the source rate. A worst case Effective Bandwidth definitions based on the traffic parameters

PCR, SCR and MBS (maximum burst size) is given in [EMW95]. Traffic with the given parameters is considered to be of ON/OFF type, which transmits at PCR during on periods of duration MBS and at rate 0 during off times, such that the mean rate is SCR.

Rate Sharing Multiplexing (RSM)

In RSM, the probability that the input rate exceeds the link rate is non-negligible. Large buffers are needed to absorb this momentary input rate excess. Such situations occur in particular in data networks (with less strict timing constraints) where connections may have large peak bit rates compared to the link rate. Rate Sharing performance heavily depends on the traffic characteristics of the input traffic. For example in the case of simple ON/OFF sources, the notion of Effective Bandwidth in REM only depends on the peak and mean rate, while for RSM also the distribution of the duration of the ON and OFF periods and the correlation between successive bursts have a significant impact. For more complicated traffic profiles, these characteristics can be very difficult to estimate.

In order to simplify CAC, also for RSM a notion of Effective Bandwidth is introduced. It can be determined on basis of the asymptotic slope of the complementary queue length distribution (see e.g. [GAN91], [EM93], etc.).

When the complementary buffer occupation distribution in a multiplexer with bandwidth C is given by $P(B > b) \approx e^{-\eta_i(c).b}$, then the Effective Bandwidth needed to obtain an overflow probability with a buffer of size B less than ε is given by $c_i = \frac{1}{\eta_i(c)}(-\log\frac{\varepsilon}{B})$. The function $\eta_i(c)$ is determined by the statistical properties of the traffic source. Remark that the above asymptotic behavior is valid for Markovian input but fails to be true for traffic with for example Long Range Dependence characteristics.

4.2 Usage/Network Parameter Control (UPC/NPC)

Once the contract between the user and the network is established and the connection is accepted, the network needs mechanisms (i) to check that the traffic is generated according to the specification and (ii) to enforce the compliance in case of violation. These actions can be performed at the User-Network Interface (UNI) and in this case it is called Usage Parameter Control (UPC) or at the Network-Node Interface (NNI), where it is referred to as Network Parameter Control (NPC). The mechanisms involved often are called policing mechanisms. It is clear that CAC and UPC/NPC are closely related and should take their decisions based on the same parameters.

UPC/NPC Requirements

The UPC/NPC is defined as the set of actions taken by the network to monitor and control traffic in terms of traffic offered and validity of the ATM connection, at the user access and the network access respectively [I371]. The main purpose is to protect network resources from malicious as well as unintentional misbehavior which can affect the QoS of other already established connections by detecting violations of negociated parameters and taking actions. In ATM networks the users can establish connections of very different bit rates over an access link of around 150 Mbit/s. This flexibility could allow a user to contract a communication at a certain bit rate and then

emit its information at a much higher rate. In order to prevent this behavior, which would degrade network performance, a policing device has to enforce the contracted bit rate at the user-network interface (and possibly at the interface between networks of different operators). This enforcement is complicated by the fact that an initially periodic cell stream is altered by the random delays affecting cells in multiplexing stages between the source and the policing device.
In general, any UPC/NPC mechanism has to comply with the following requirements:

- the ability to detect any illegal traffic situation
- the ability to determine if the traffic is compliant
- fast reaction to parameter violations
- transparency to compliant traffic
- easy to implement.

A UPC/NPC mechanism has to decide whether a random cell flow is conforming or not. Therefore, such a mechanism can not be perfect and, even if the user respects its traffic contract, a certain number of cells will be erroneously detected as non-conforming. This error rate will have to be maintained very low (for example 10^{-10} if we want to ensure a cell loss rate of 10^{-9}). An other kind of error will occur when the mechanism does not detect any violating cells while the traffic is non conforming. As explained, as the observed flow is random the mechanism cannot always distinguish between an illegal traffic situation and one due to the CDV introduced by the network.

Performance Parameters

Two performance measures are widely used to evaluate and compare the control mechanisms:

- response time: the time needed by the mechanism to detect a contract violation
- transparency: the accuracy with which the UPC initiates appropriate control actions on a non-compliant connection and avoids inappropriate control actions on a compliant connection.

There is a trade-off between these two parameters: if we want to be very accurate the response time will increase. When a policing mechanism has to observe at least N cells (equal to the number of credits) before being able to detect a violating cell, then a large value of N will slow down the reaction time.

UPC Location

The policing function is part of the public network, but it should be located as close as possible to the user. Therefore, the UPC function is located where the Virtual Channel Connections (VCC) or Virtual Path Connections (VPC) are terminated within the network. This implies that UPC is performed before the first switching activity takes place.

UPC Actions at Cell Level

A UPC mechanisms may perform the following actions at the cell level: cell passing, cell re-scheduling, cell tagging and cell discarding. Cell passing and cell re-scheduling are performed on cells which are identified by a UPC/NPC as compliant. Cell re-scheduling is performed when traffic shaping and UPC are combined. Cell tagging

and cell discarding are performed on cells which are identified by a UPC/NPC as non-compliant. Cell tagging operates on CLP=0 cells only by overwriting the CLP bit to 1.

Policing Mechanisms

Some authors (see e.g. [BOY92a], [GUI92]) distinguish between two classes of control mechanisms: the so-called "pick-up" mechanisms and the ones which shape the traffic.
A pick-up mechanism observes a cell flow and detects the exceeding cells. Therefore, or else the cells pass transparently through the policing device, or else they are detected as violating the contract and they are dropped or tagged. A shaper, in general, modifies the traffic even if it is non-conforming.
Several pick-up mechanisms have been proposed and evaluated. These mechanisms can be classify as follow:

- Window Mechanisms such as:
  - The Moving Window Mechanism
  - The Jumping Window Mechanism
  - The Triggered Jumping Window Mechanism
  - The Exponentially Weighted Moving Average Mechanism
- The Leaky Bucket Mechanism
- The Virtual Scheduling Mechanism

We restrict this discussion to the most prominent scheme, the Leaky Bucket Mechanism.

The Leaky Bucket Mechanism (LB)

The LB mechanism which was introduced by Turner in [TUR96] is probably the most well known policing mechanism and it has been thoroughly evaluated in many papers (e. g. [RAG91], [NIE90]...)
The LB mechanism uses three control parameters:

- a threshold value (N)
- a decrementing value (d)
- a decrementing frequency (1/T)

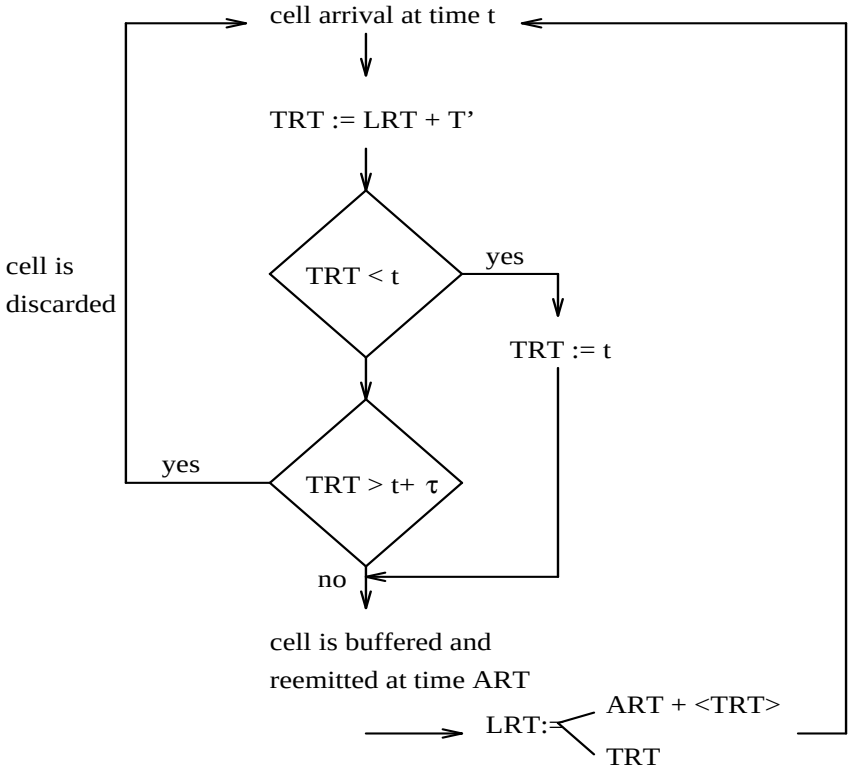
The LB increments its counter value for each arrival if it is below the maximum N and discards the cell if the counter value is N upon arrival. The counter value is decremented by d every T time units if it is above 0 (Figure x).

The policing bit rate is in that case $\lambda = \frac{d}{T} \times 48 \times 8$ bits/sec.

If d>1, the decrementing frequency T can be increased and the policing algorithm can operate more slowly. This can be advantageous in the case of monitoring high bit rates. Until the counter reaches the threshold value N, the LB allows short exceeding of the policing bit rate. The LB can let pass $n_{\max}$ cells in a full-rate burst at 155 Mbit/s, with [NIE90]: $n_{\max} = N + kd$, with $k = 1 + \frac{N + d}{\frac{T}{2.83\mu s} - d}$.

Of course, the policing algorithm should limit the length of the bursts which means that N should be as small as possible. On the other hand, cells may not be discarded due to jitter introduced by the network if the user does not violate the negociated

connection parameters. The leaky bucket can be dimensioned as a function of the CDV introduced by the network using the well known formula proposed in [NIE90] $1+d+\delta\times r\times(8/48)<N$, with r the bit rate of the policed CBR connection and δ the CDV.



TRT: cell Theoretical Re-emission Time

ART: cell Actual Re-emission Time

LRT: cell Last Re-emission Time

Figure 4 : The Spacer-Controller

4.3 Traffic Shaping

Traffic shaping is a traffic control mechanism which alters the characteristics of a cell stream. An important class of traffic shaping mechanisms are the *Spacers*. By spacing the cells of a connection in time, the peak bit rate may be reduced, the Cell Delay Variation (CDV) may be controlled or the burst duration may be limited. Traffic shaping may be performed on different locations.

Traffic shaping in the Customer's Premises Network (*CPN*).

As mentioned before, the UPC function uses the traffic descriptor to check whether the cells stream offered to the network is conforming the contract. In order to enforce a source to be conform to the traffic contract, its cell stream may be shaped in the CPN before entering the network to obtain the required traffic characteristics.

Traffic shaping in the ATM network.

Passage through multiplexers and switches may alter the characteristics of a traffic stream considerably. In particular due to the queueing delays, the stream is jittered leading to cell clumping and dispersion of cells. These phenomena imply an important decrease of network utilization to obtain a given QoS. Therefore, traffic shaping

within the network is applied to change the traffic characteristics such that a higher utilization is achieved.

The aim of a traffic shaper is to alter the traffic characteristics of a stream of cells of a VCC or a VPC to achieve a desired modification of those traffic characteristics. In order to do that, a traffic shaping mechanism can perform the following actions:

- reduce the peak cell rate
- limit the burst length
- reduce CDV by suitably spacing cells in time
- use different queue service schemes.

A traffic shaper can re-shape the traffic at the entrance of the network and allocate resources in order to respect both the CDV and the propagation delay allocated to the network. A possible alternative is to dimension the network in order to accommodate the input CDV and provide for a shaper at the output. A final possibility could also be to dimension the network in order both to accommodate the input CDV and comply with the output CDV without any shaping function.

Scheduling disciplines

Several queue service schemes have been proposed in order to be able to provide multiple QoS.

Generalized Round Robin (GRR)

The GRR [ROJ94] distinguishes an individual queue for each multiplexed connection. In the classical round robin discipline each queue is visited cyclically with at most one customer being served at each visit. The generalization consists of allowing the visit frequency to be different for each queue. The visit frequency would be determined by a bandwidth reservation parameter which, e.g., could be for a high speed data connection the sustainable cell rate, for a CBR connection the peak rate and for a low peak VBR connection some intermediate "equivalent rate". Direct implementation of GRR in ATM is apparently still quite complicated. A "queueing engine" allowing GRR is described in [KMI92].

Fair Queueing (FQ)

FQ [DKS89] defines a separate FCFS queue for each connection and if k of these queues are currently not empty, then each non empty queue receives 1/k-th of the link bandwidth. Different bandwidth demands can be expressed using relative weights [CSZ92].

Virtual Clock (VC)

In this scheme [ZHA91], the cells of a given stream i with bandwidth allocation Λ_i are allocated a time stamp on arrival and all cells in the multiplex queue are served in increasing order of time stamp. The time stamp of cell number n+1 is equal to the greatest of the time of cell n plus the maximum intercell interval $\frac{1}{\Lambda_i}$ and the current time. A cell is served as soon it reaches the head of the queue.

Virtual Spacing (VS)

The VS [ROJ94] realizes the GRR queue discipline. As in the Virtual Clock algorithm, cells destined to a given output multiplex are attributed a time stamp which determines their order of service. However, only one cell per connection is stamped at any time, the stamp being attributed to a new cell only after the previous cell has been transmitted. The cells of any given connection are stored in a FIFO queue: when the first cell in the queue is transmitted at a certain time t, the next cell is attributed the time stamp $t+\frac{1}{\Lambda}$ and will be served as soon as no other cell for the same multiplex has a time stamp of smaller value. If, when a cell is transmitted, no further cells of the same connection are queued, the next cell to arrive will be attributed a time stamp equal to the maximum of the current time and the value $t+\frac{1}{\Lambda}$. If the VS would determine the new time stamp from the previous time stamp rather than the actual transmission time, we would have the VC algorithm.

Jitter Earliest-Due-Date
After service, a cell is stamped with the difference between its deadline and its actual finish time. The next switch will hold this packet for an extra amount of time equal to the calculated difference [VER91].

Stop and Go Queueing
Stop and Go Queueing [GOL90] consists of imposing a synchronized frame structure on the network guaranteeing the availability of transmission slots at the appropriate times for periodically arriving cell streams with real time constraints.

Spacing Algorithms

Conformance to the traffic contract at the network entry point does not imply that the traffic offered by the connection respects the negociated peak emission interval. It has been shown that the pick-up policing functions previously described do not prevent clusters of cells from entering the network and therefore cannot protect the network from congestion under all conditions. This is due to the jitter tolerance which has to be introduced in order to accommodate for the random delays introduced on the cell flow in successive multiplexing stages. The policing function is not able to decide whether short bursts that violate the specified peak cell rate are caused by delay jitter or by misbehaving customers. Pick-up policing devices therefore let them pass transparently through. The problems introduced by these clusters of cells which would pass can be avoided if the policing function not only discards excess cells but also delays cells so that their interdeparture times from the policing device between cells of one connection are never below a minimal value which is chosen according to the negociated peak cell rate. The Spacing Function is intended to guarantee that an incoming traffic conforms to the negociated peak cell rate. Several implementations of such devices which combine a pick-up policing function with a spacer have been proposed in the literature [BOY92a], [WAL91]. We briefly described here the Spacer-Controller proposed in [BOY92a] which is composed of:

- a control block which performs a strict enforcement of the peak bit rate value while accommodating for a standardized CDV altering the connection upstream
- a spacing block which ensures a minimum spacing between two consecutive cells.

In the Spacer-Controller, cell flows are policed according to a Virtual Scheduling Algorithm. In addition, a Cell Spacing Algorithm is used to retrieve, as far as possible, the Peak Emission Period corresponding to the negociated peak cell rate. In Figure 4, T is the peak emission period of the connection negociated in the traffic contract, τ is the CDV tolerance. The variable called TRT is the cell Theoretical Reemission Time. If a cell was to be spaced, it should not depart from the UPC/NPC before time TRT. When TRT is larger than the current time t, the policing function is actually receiving a cell clump, in principle due to CDV. The policing function virtually smoothes out the cell clump until the cell re-emission schedule goes beyond a limit τ ahead of the current time t. Before this limit, the cell is declared as conforming to the traffic contract and TRT is stored in the connection context memory. When this limit is reached, the cell is declared as non-conforming and subject to a policing action. In this case, TRT is not updated in the context memory. When TRT is smaller than the current time t, the incoming cell has been delayed so much that re-emission schedule is becoming late. This cell is probably the first of a cell clump. In order to avoid replication of the clump, TRT is set to the current time t. The spacing function has to schedule the reemission of an incoming cell on the link at the output of the UPC/NPC. TRT is rounded to the next upper discrete time value which is the cell *Actual Reemission Time* ART. However, it can happen that reemission of a cell belonging to another connection has been already scheduled at the same time slot. Therefore, the spacing function has to find the first available time slot after TRT to set ART.

4.4 Priority Control and Selective Cell Discarding

The header of each ATM cell contains a Cell Loss Priority (CLP) bit. This bit is used to indicate a loss priority. The network may decide to selectively discard cells with low priority in favour of high priority cells. Priority marking can be performed either on a connection basis or on a cell basis. We give two examples of potential use of priority control in ATM networks.

(i) Different classes of QoS: By using the CLP bit on a connection basis, the network may distinguish two different classes of QoS : traffic for which CLP=0 and traffic for which CLP=1. In this case the network may guarantee different cell loss rates according to the class a connection belongs to and it must provide selective discard mechanisms in order to handle the different classes. Examples of such mechanisms are pushout, partial buffer sharing [KRO90], [SUM88]. The increase in complexity of network elements due to these mechanisms may be compensated by the possible increase in accepted load in the network due to the existence of different classes of QoS with different cell loss ratio guarantees.

(ii) Cell tagging: When the UPC function detects a cell which violates the contract it may discard the cell or tag it as non-conforming (for a detailed discussion see Section 4.1.5). In the later case, the CLP bit may be used to indicate whether a cell is conforming or not. As soon as congestion occurs in the network, the CLP bit may then be used to selectively discard non-conforming cells.

4.5 Fast Resource Management

Statistical multiplexing may lead to a more efficient use of the network resources at the expense of additional traffic control functions. A typical example of this principle may be found in Fast Resource Management, where control is performed on the time scale of the round-trip propagation delay of an ATM connection. Let us give an example of such a control mechanism.
In order to obtain a statistical multiplexing gain, the network should not allocate the peak bit rate for the whole duration of the connection for Variable Bit Rate (VBR) traffic. In addition, typical services generating VBR traffic, e.g. data services, tolerate a certain delay. These observations lead to the notion of the *Fast Reservation Protocol* (see [BOY92b]). The idea is to allocate the necessary bandwidth to a connection for the duration of a burst only. By means of Reservation Request Cells, a source indicates the desire to increase its bit rate. Two variants exist : *Fast Reservation Protocol with Delayed Transmission* (FRP/DT), where the source waits for an acknowledgement from the network (by means of a Reservation Accepted Cell) before increasing its activity and the *Fast Reservation Protocol with Immediate Transmission* (FRP/IT), where the burst is transmitted immediately after the request cell. In the later case, the whole burst is discarded in case the reservation fails. These schemes implement the ABT Service Category.

4.6 Network Resource Management

Network Resource Management (NRM) is a subset of Traffic and Congestion Control (TCC) functions related to resource configuration and allocation. The main networking technique is the use of VPCs. Managing these virtual path connections may involve [BUR90], [BUR91] allocating capacity based on anticipated demand, rerouting traffic in times of congestion, changing allocations of capacity to cater for changing demand. By reserving capacity on VPCs, the processing required to establish individual VCCs is reduced: individual VCCs can be established by making simple connection admission decisions at nodes where VPCs are terminated. VPCs can be used to [I371] :

- simplify CAC
- implement a form of priority control by segregating traffic types requiring different QoS
- aggregate user-to-user services such that the UPC can be applied to the traffic aggregation.
- efficiently distribute messages for the operation of traffic control schemes.

This use can lead to the following advantages: a reduced load on control equipment; lower call establishment delays; additional means of providing service protection to improve network availability; an additional means of controlling network congestion.

5. Congestion Control for Best Effort Services

5.1 ABR Flow Control Schemes

The ATM Forum has proposed a number of congestion control mechanisms for the ABR service class. The two most important classes of proposals are the credit-based schemes and the rate-based schemes. Eventually the rate-based scheme was adopted as solution for congestion control of the ABR service class. The congestion scheme described in the ATM Forum specifications [ATM95] is a rate-based, closed-loop, per-connection control which uses the feedback information from the network to regulate the rate at which the sources transmit cells. The transmission rate of each connection is controlled by means of special control cells called resource management

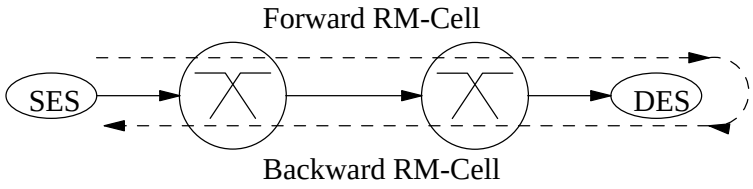


Figure 5 : ABR flow control

(RM) cells. RM-cells flow from the source end system (SES) to the destination end system (DES) and return along the same path carrying congestion information (Figure 5). Depending on the congestion information received in the RM-cell, the SES increases or decreases its transmission rate. The standard specifies the source and destination behavior and several methods that a switch can implement to control congestion.

SES and DES behavior

At the connection set up the source negotiates the maximum and minimum rate at which it may transmit (PCR and MCR) ; the initial cell rate (ICR) at which it may start transmitting ; the number of cells per RM-cell (Nrm) ; the rate increase factor (RIF) and the rate decrease factor (RDF). The flow chart in Figure 6 shows the source behavior. The SES starts transmitting with the agreed ICR. Each Nrm-1 data-cell transmissions, the SES sends an RM-cell with the following fields : Explicit Rate (ER) set to PCR ; Current Cell Rate (CCR) set to the Allowed Cell Rate (ACR) of the source ; Congestion Indication Bit set to 0 (no congestion) ; No Increase (NI) bit set to 0 (no increase) and Direction (DIR) bit set to forward. The ACR value establishes an upper bound to the transmission rate of the source. The source may transmit at the ACR while not becoming idle or rate-limited.
The cells are received by the DES which must store the Explicit Forward Congestion Indication Bit (EFCI) of the last Data-Cell received. On receiving a forward RM-cell it must change the CI bit to congested state depending on the EFCI bit stored, change the DIR to backward and send the RM-cell back to the SES along the same path.
On receiving a backward RM-cell the SES adjusts the ACR. When a backward RM-cell is received with CI = 0 and NI = 0, the SES is allowed to increase its rate (ACR) by no more than RIF*PCR. On receiving an RM-cell with CI = 1, the SES must decrease the ACR by at least RDF*ACR. Finally the ACR must be set at most to the ER field. The ACR cannot be reduced below the MCR or increased above the PCR. The actions marked as “Rescheduling option” are an optional behavior which allows to reschedule the transmission time of a cell in order to take advantage of an increase in the ACR. The actions marked as “ADTF adjustment” (ACR Decrease Time Factor) are used to control the ACR during the idle periods of the source. After such a period

the source could start transmitting at the full ACR, resulting in a harm for the network if the last computed ACR was too high. The ADTF adjustment consists of measuring the elapsed time between two forward RM-cell transmissions. If this time is higher

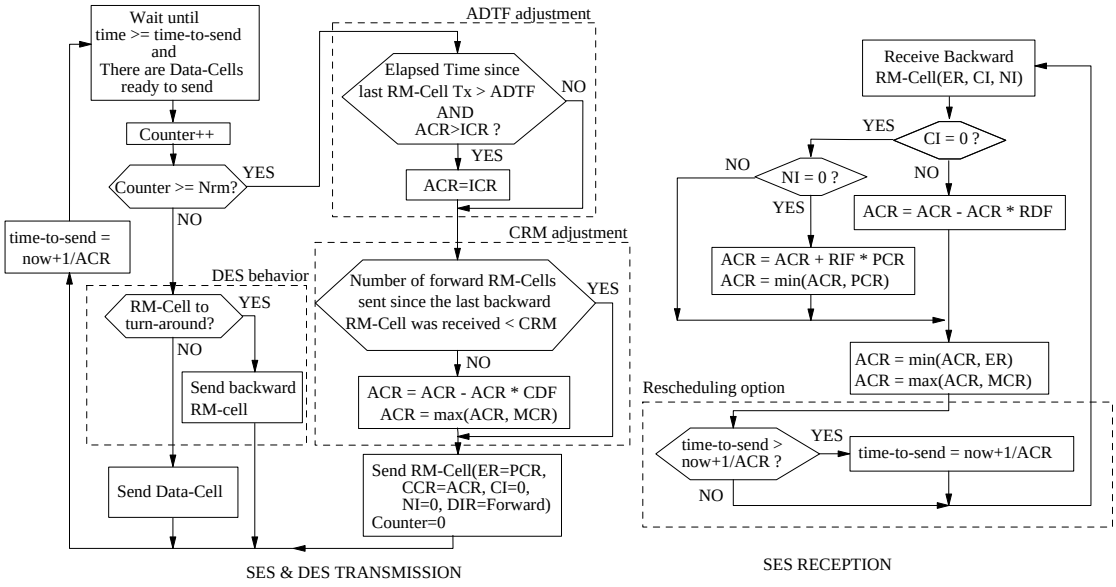


Figure 6 : Source Behavior

than the ADTF, the ACR is reduced down to the ICR. We note that if a source becomes rate-limited but not idle, it could also start transmitting at the full ACR and the ADTF adjustment would not work. To avoid this the ATM Forum establishes the so called “use-it-or-lose-it” optional behavior which consists of reducing the ACR in order to maintain it reasonably close to the transmission rate of the source.
The actions marked as “CRM adjustment” constrain the source to reduce the ACR in case of absence of backward RM-cells reception. This condition could be caused by a heavy congestion state of the network. If the number of forward RM-cell transmissions since the last backward RM-cell reception is higher or equal to CRM, the ACR must be reduced by, at least, ACR*CDF.

ABR Switch Mechanisms

A switch shall implement at least one of the following methods to control congestion : set the EFCI bit of the data cells ; set the CI or NI bit in forward and/or backward RM-cells ; reduce the explicit rate (ER) field of forward and/or backward RM-cells. The switches that set the EFCI or CI bit to indicate a congestion state are known as binary switches. Switches that modify the ER field are called ER switches.
Several switch mechanisms compatible with the ATM Forum specifications have been proposed. They differ on the congestion monitoring criteria and the feedback mechanism used. We describe three of them which are well known to show the different degrees of performance and complexity that can be achieved.

EFCI Switch

The simplest switch mechanism [YIN94] marks the EFCI bit in data cell headers when congestion is detected. The switch monitors its queue length and detects

congestion when it exceeds a given threshold. The feedback delay can be reduced by setting CI = 1 of backward RM cells during the congested state instead of setting the EFCI.
The main drawback of this switch mechanism is its lack of fairness. Depending on the network topology, some connections can have an unfair access to the available bandwidth. For example, RM-cells of a VC going through a higher number of congested links will be set to congested more often than those of VCs going through fewer congested links. This undesirable effect (known as the "beat down problem") will result in a lower rate for such VCs.

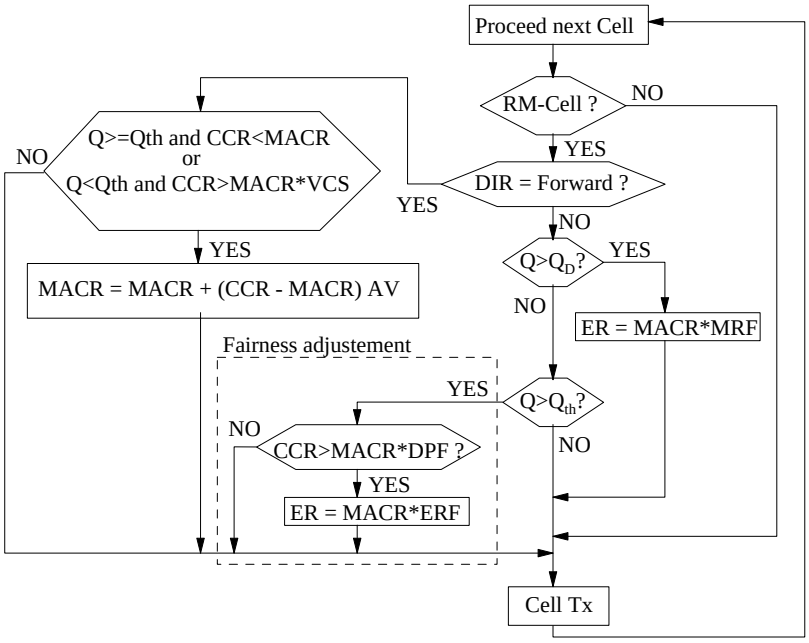


Figure 7 : EPRCA Switch

EPRCA Switch

The Enhanced Proportional Rate Control Algorithm (EPRCA) [ROL94] is an enhanced version of the original rate-control algorithm. The switch computes an heuristic approximation of a fair rate, equal to the link capacity minis the capacity of the constrained VCs over the non constrained VCs (max-min criterium). The fair rate (MACR in the figure) is computed during the uncongested periods as an exponential average (MACR = MACR + (CCR - MACR) AV) over all the VCs whose CCR is larger than MACR*VCS. AV is the averaging factor and VCS is a VC separator used to distinguish between VCs constrained by the switch and otherwise constrained VCs (see Figure 7). To avoid the "beat down problem", the switch just reduces during congested periods the ER field of the backward RM-cells with a CCR greater than MACR*DPF. The ER is reduced to MACR*ERF. The Down Pressure Factor (DPF) is used to cause the rate setting control when the ACR reaches a value slightly lower than the MACR. The Explicit Reduction Factor (ERF) is used to set the explicit rates slightly below MACR so that the switch will stay uncongested. The switch is considered congested when the queue length (Q) is greater than a threshold (Qth). If Q is greater than another threshold QD, the switch is considered very congested and ER is reduced in all backward RM-cells to MACR*MRF (MRF is a major reduction factor).

ERICA Switch

The objective of the Explicit Rate Indication for Congestion Avoidance (ERICA) algorithm [JAI95] is to keep the queue length low and achieve max-min fairness. The switch mechanism is described in the flow chart of Figure 8. The main difference with the previous switch mechanisms is the detection of the congestion state. In the previous mechanisms this detection is based on a queue length threshold. In the ERICA proposal, the switches measure the input rate (IR) and compare it with a target cell rate (TCR, set to 85-95% of the link bandwidth) to compute the overload factor OF = IR/TCR. The ER field of backward RM-cells is then reduced by the OF in order to avoid the congestion state.

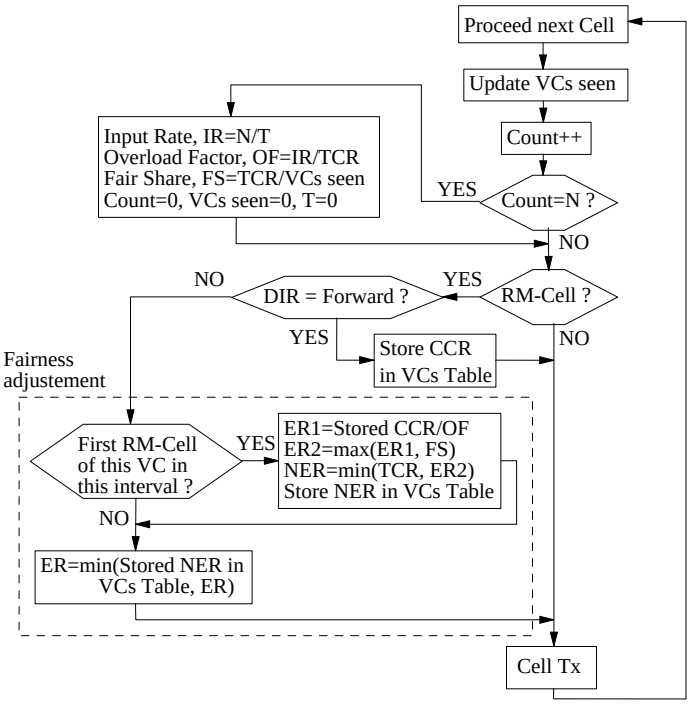


Figure 8 : ERICA Switch

To compute the IR, the switch measures the time T until N cells arrive. Then it computes IR = N/T and starts another measuring interval. During each measuring interval, the switch also counts the number of active VCs in order to compute the fair share (FS) as FS = TCR/Number of VCs seen during the measuring interval. When receiving a backward RM-cell the switch computes the explicit rate ER2 based on load and fairness (ER2 = max(CCR/OF, FS)) and stores the value NER = min(TCR, ER2). If the ER field of the cell is higher than NER, the field is replaced by the computed value.

To reduce the feedback delay the ER computation uses the CCR seen in the last forward RM-cell of the same VC. Therefore, this value must be stored in a VC table when a forward RM-cell is received.

To compute the ER the switch uses the IR and FS values computed in the previous interval. In order to avoid oscillations the proposal establishes that all the backward RM-cells of a given VC seen during the same measuring interval must use the same NER value. The switch must then store the NER computed when the first RM-cell is seen during a measuring interval, and keep an indication that a backward RM-cell of that VC has been seen in the current measuring interval.

Comparison of the switch mechanisms

EFCI is the simplest switch mechanism. It only monitors the queue length and marks backward RM-cells when higher than a threshold. However it has been shown that high queue length can be reached and fairness cannot be guaranteed.

The EPRCA switch mechanism, which computes an average ACR reading the CCR field of RM-cells and modifying the ER field of backward RM-cells, achieves a better performance in terms of link utilisation, queue length and fairness than the EFCI switch [EXP96].

The ERICA switch mechanism is the most complex. It requires measuring the input rate of each buffer and accessing to a VC table each time a forward or a backward RM-cell is received. However it achieves a high degree of fairness and a tight queue length control. Another advantage compared to the EPRCA is the reduced number of parameters to be tuned (the target utilisation and the measuring interval in cells).

ABR Conformance Definition and Policing

To control ABR sources and to check whether or not they respond to the feedback information, a conformance definition is introduced in standardisation. An example of a conformance definition for an ABR connection based on the Dynamic Generic Cell Rate Algorithm (DGCRA) has been defined by ITU-T [I371] and the ATM Forum [ATM95]. The conformance definition is a part of the traffic contract which defines a reference algorithm used to define whether the cells passing a measuring point located at the UNI are conforming or not. A network operator may use a Usage Parameter Control (UPC) which, based on the conformance definition, defines whether a connection is compliant or not. The UPC may mark or discard non-conforming cells.

What is new for ABR connections is the variable lag which exists between the moment a rate change is communicated to the source and the time this change is observed at the interface. This can be seen in Figure 9. Forward RM-cells generated by the SES are inserted in the data flow and contain a value for the ER at which the source would like to transmit. These RM-cells are looped back by the DES to the SES. Nodes in between can access the ER field and lower the ER in case of congestion. Depending on the distance between source and interface and on the

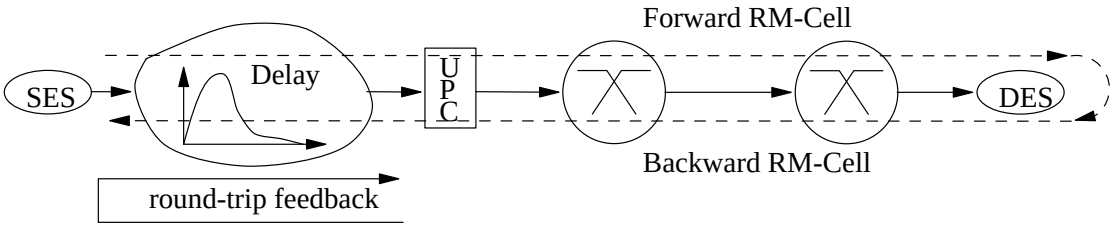


Figure 9 : Policing ABR Traffic

background traffic, it will take a variable time before the RM cells arrive at the interface. In order to have the most recent value of the ER, the policing function must also check the flow of RM cells in the backward direction of the connection. This is also new compared to policing CBR and VBR traffic where only the direction from source to destination has to be monitored.

In order to cope with the variable lag between source and interface two time constants τ_2 and τ_3 are introduced which are respectively an upper bound and a lower bound of this round trip delay.

Because of the variable available bandwidth to which the source must adapt itself, the source traffic characteristics will be altered during the lifetime of the connection. The GCRA (used for the conformance definition for CBR and VBR) is a static algorithm in the sense that its two parameters, the increment value I (e.g. the inverse of the PCR) and the limit L (e.g. the CDV tolerance) are not allowed to vary (without re-negotiation of the traffic contract). The DGCRA has the same two parameters as the GCRA but the increment parameter is allowed to vary (without re-negotiation of the traffic contract) between the inverse of the PCR and the inverse of the MCR. The computation of this varying increment is not an easy task because a rate change conveyed by a backward RM-cell received at the interface at a given time may be applied to the forward cell flow after a variable delay t (τ_3 =< t =< τ_2). To be on the safe side, the DGCRA schedules rate increases conveyed in the backward RM-cell flow after a delay of τ_3 and rate decreases after a delay of τ_2.
Two algorithms have been proposed to compute the variable increment. Algorithm "A" provides the tightest conformance according to the delay bounds but is rather complex to implement and a simpler algorithm "B" has been defined which is much less accurate [CER96].

5.2 Intelligent Packet Discard Schemes for UBR

The absence of any QoS guarantee for the UBR Service Category may lead to a low throughput. The loss of a single cell in the network of an AAL5 PDU frame inevitably leads to the discard of the whole frame at the destination. Delivering such a corrupted frame leads to an important waste of network resources and may drop the effective throughput drastically. This is in particular true in a broadband network where the high data rate and the long distances force the retransmit/recovery mechanisms to retransmit a high number of frames already sent out since the corrupted cell (depending on the window size). Another disadvantage of the absence of any traffic and congestion control mechanism in the UBR Service Category is the lack of fairness between the different connections that share the bandwidth using UBR. Indeed, connections which lose cells will be forced by the transport protocol (e.g. TCP) to slow down their transmission rate, allowing other connections to use more bandwidth and buffer space. These observations have lead to the introduction of intelligent packet discarding mechanisms. In what follows, we discuss Partial Packet Discard, Early Packet Discard, Per VC Accounting and Per VC Queueing.

Partial Packet Discard

Partial Packet Discard (PPD) (see e.g. RF95]) (also called Packet Tail Discard [TUR96], Partial Packet/Frame Drop [LNO96], Drop Tail [FCH94], Tail Dropping [KKT96]) is a packet discarding scheme which drops the remainder of a frame, apart from the End of Message (EOM) cell, as soon a a cell loss occurs. The EOM cell is needed by the destination end station to delineate the beginning of a new frame. The scheme is called partial, as only those cells of a frame are dropped arriving after the occurrence of a loss. There is no de-queueing of already accepted cells. The scheme recognizes the beginning of a new frame by inspecting the Payload Type Identifier (PTI) field in the header of incoming cells. The PTI contains an indication of the EOM.

Early Packet Discard

This scheme enforces a network element to drop an entire frame (i.e. all its cells) when the first cell of that frame arrives at a buffer which exceeds a predefined threshold. There are several methods to set the threshold. In [RF95], a fixed threshold is proposed, while in [KKT96] a variable threshold is used based on the number of already accepted frames. The Random Early Detection (RED) algorithm proposed in [LNO96] uses the observed average queue size to define a probability by which a frame is dropped in case of congestion.
Although the throughput may be increased considerably using EPD, the fairness problem still remains, due to the fact that EDP does not consider the number of already accepted frames per VC when dropping a newly arriving frame. The following methods try to solve this problem.

Per VC Accounting

In a per VC accounting scheme, frames are discarded not only on basis of the buffer occupation level, but also based on the origin of the frames in the buffer. Apart from the buffer occupation threshold, the scheme computes a Fair Buffer Share (FBS) defined to be FBS = K $\times$ (Total Buffer Occupation / number of active VCs).
The constant K is a factor for the buffer occupation and is chosen as $1 < K < 2$. As soon as the threshold of the buffer occupation is reached, then all frames of overloading connections (i.e. connections which use more than the a number FBS of buffer places) are dropped. Frames of non-overloading connections are still accepted. This mechanism has the advantage that it co-operates better with the transport protocol (e.g. FTP). Indeed, when a cell of a connection is lost, the transport protocol will enforce this connection to drop its frame generation rate, implying a lower cell arrival rate at the buffer, resulting in a highly probable situation where the buffer occupation of that connection is below the FBS. Hence, a connection that has experienced a cell loss has a higher probability of having its next frame(s) accepted, even if the congestion is persisting.

Per VC Queueing

In this scheme, the total buffer capacity is subdivided in a number of logical queues, one for each active connection. These queues are served according to a Round Robin Scheme, implying fairness even during intervals without congestion. The mechanisms during periods of congestion are similar to Per VC Accounting, i.e. a buffer occupation threshold is chosen which indicates congestion and as soon as this threshold is surpassed, frames are dropped of those connections with buffer occupancy higher than a Fair Buffer Share. Clearly this mechanism is more complex than the other ones, but performance studies have shown its superiority with respect to effective throughput and fairness (see e.g. [JGK96]).

5.3 Guaranteed Frame Rate

The mapping of the GFR frame level guarantee onto an appropriate cell level guarantee is achieved based on the identification of the frames to which the service

guarantees apply. This can be done by using a modified GCRA(1/MCR, Burst Tolerance(MBS)+CDVT), where MBS = 2*CPCS-SDU size (in cells).
Two main implementations to support GFR have been proposed and are currently being discussed [ATM97]:
(i) *GFR implementation using Weighted Fair Queuing (WFQ) and per-VC accounting*
This implementation serves an individual VC at a rate of at least MCR using a WFQ scheduler. The buffer management is based on a per-VC accounting so that each ATM connection can have its own part of available bandwidth and buffer.
(ii) *GFR implementation using tagging and FIFO queue*
In that case the cell rate guarantee of GFR cannot be provided by the service discipline and a tagging function is needed to identify cells eligible for service guarantee. The modified GCRA(1/MCR, Burst Tolerance(MBS)+CDVT) is used to determine which cells to tag.
As GFR looks promising with respect to the efficient transport of TCP traffic, the performance of TCP over GFR is thoroughly being investigated. Early simulations indicate that the GFR implementation based on FIFO queuing and tagging is not able to provide the cell rate guarantee to a TCP source while the other implementation allows to provide satisfactory performance of TCP over GFR.

6. Conclusions

In this paper, an overview is presented of the various Service Categories and Transfer Capabilities in ATM networks, together with the traffic control and congestion control mechanisms that support the QoS guarantees offered by these categories. Today, a number of concepts and mechanisms are specified or even standardized, such as the traffic parameter definition using the GCRA, the leaky bucket algorithm for UPC, the rate based flow control scheme for ABR traffic, etc. Other mechanisms, such as CAC, are system dependent and remain also today a topic of intensive research and competition in commercial ATM products. Another development that has heavily influenced the traffic management architecture is the world-wide use of Internet and the related protocols and applications. In particular the possibility to offer QoS guarantees is an important topic in the discussion on Internet and ATM. Studies and experiments have shown that, in order to carry Internet traffic in an efficient and economical way over an ATM network, new control mechanisms that operate on layers above ATM are needed (for example to guarantee the goodput of AAL5 PDUs). The current development of the GFR Service Category illustrates this trend.
In spite of the fact that already a lot of effort has been put in Traffic Management research and development activities, there remain many questions unanswered and further research effort is needed in this area.

References

[ATM95] ATM Forum Technical Committee Traffic Management Working Group, *"ATM Forum Traffic Management Specification Version 4.0"*, ATM Forum, October 1997

[ATM97] Traffic Management Working Group, *"Guaranteed Frame Rate Service (GFR)"*, item 96-003 of the living list, July 1997.
[BOY92a] P.E. Boyer, F.M. Guillemin, M.J. Servel and J-P. Coudreuse, *"Spacing Cells Protects and Enhances Utilization of ATM Network Links"*, IEEE Network Magazine, Vol. 6, No. 5, September 1992.
[BOY92b] P.E. Boyer, D. Tranchier, *"A Reservation Principle with Applications to the ATM Traffic Control"*, Computer Networks and ISDN Systems, 24, North Holland, 1992, pp. 321-334
[BUR90] J. Burgin, *"Dynamic Capacity Management in the BISDN"*, Int. Journal of Digital and Analog Communication Systems, Vol. 3, pp.161-165, 1990.
[BUR91] J. Burgin and D. Dorman, *"Broadband ISDN Resource Management: The Role of Virtual Paths"*, IEEE Comm. Mag., Vol. 29, No. 10, pp. 44-48, 1991.
[BUT91] M. Butto, E. Cavallero, A. Tonietti, *"Effectiveness of the "Leaky Bucket" Policing Mechanism in ATM Networks"*, IEEE JSAC, Vol. 9, No. 3, pp. 335-342, 1991.
[CER96] L. Cerda, O. Casals, *"Improvements and Performance Study of the Conformance Definition for the ABR Service in ATM Networks"*, ITC Specialist Seminar on Control in Communications, Lund, Sweden, September1996.
[COST242] J. Roberts, U. Mocci and J. Virtamo (Eds), *"Broadband Network Teletraffic"*, Final Report of Action COST 242, Springer Verlag 1996
[CSZ 92] D.D. Clark, S. Shenker and L. Zhang, *"Supporting Real Time Applications in an Integrated Services Packet Network: Architecture and Mechanisms"*, ACM SIGCOM'92, 1992.
[DKS89] A. Demers, S. Keshav, S. Shenker, *"Analysis and Simulation of a Fair Queueing Algorithm"*, ACM SIGCOM'89, 1989.
[EM93] A. Elwalid and D. Mitra, *"Effective bandwidth of general Markovian traffic sources and admission control of high speed networks"*, IEEE/ACM Trans Networking, 1, June 1993
[EMW95] A. Elwalid, D. Mitra and R. Wentworth, *"A new approach to allocating buffers and bandwidth to heterogeneous regulated traffic in an ATM node"*, IEEE J. Selected Areas in Comm., 13(6), August 1995, p.1115-1128
[EXP96] Deliverable 6 of the ACTS Project AC094 EXPERT, *"Specification of Integrated Traffic Control Architecture"*, September 1996.
[FCH94] C. Fang, H. Chen and J. Hutchins, *"A simulation study of TCP performance in ATM networks"*, Proceedings of IEEE INFOCOM '94, Vol.2, San Francisco, 1994, p.1217-1223
[GAN91] R. Guerin, H. Ahmadi and M. Naghshineh, *"Equivalent capacity and its application to bandwidth allocation in high speed networks"*, IEEE J. Selected Areas in Comm., 9, 1991, p.968-981
[GIL91] H. Gilbert, O. Aboul-Magd, V. Phung, *"Developing a Cohesive Traffic Management Strategy for ATM Networks"*, IEEE Comm. Mag., Vol. 29, No. 10, 1991.
[GOL90] S. Golestani, *"Congestion-free Transmission of Real-Time Traffic in Packet Networks"*, Proc. IEEE Infocom´90, pp. 527-542, San Francisco, CA, June 1990.
[GUE96] R. Guerin, J. Heinanen, *"UBR+ Service Category Definition"*, ATM Forum contribution No. 96-1598, December 1996.
[GUI92] F. Guillemin, P. Boyer and L. Romoeuf, *"The spacer-controller : architecture and first assessments"*, Proc. IFIP Workshop on Broadband Communications, Estoril, Portugal, 1992.

[I371] CCITT Draft Recommendation I.371 (now ITU-T I.371), *"Traffic Control and Resource Management in B-ISDN"*, Melbourne, Dec. 1991.
[JAI95] R. Jain et al., *"A Sample Switch Algorithm"*, ATM Forum contribution No. 95-0178R1, February 1995.
[JGK96] R. Jain, R. Goyal, S. Kalyanaraman, S. Fahmy and F. Lu, *"TCP/IP over UBR"*, ATM Forum contribution 96-0179
[KEL91] F. Kelly, *"Effective bandwidths at multi-class queues"*, Queueing Systems, 9, 1991, p.4-15
[KKT96] K. Kawahara, K. Kitajima, T. Takine and Y. Oie, *"Performance evaluation of selective cell discard schemes in ATM networks"*, Proceedings of IEEE INFOCOM '96, Vol.3, San Francisco, March 1996, p.1054-1061
[KMI92] C.R. Kalmanek, S.P. Morgan, R. C. Restrick III, *"A High-Performance Engine for ATM networks"*, ISS'92, 1992.
[KRO90] H. Kroner, *"Comparative Performance Study of Space Priority Mechanisms for ATM Channels"*, IEEE Infocom'90, San Francisco, June 1990.
[LNO96] T. Lakshman, A. Neidhardt and T. Ott, *"The drop from front strategy in TCP over ATM"*, Proceedings of IEEE INFOCOM '96, Vol.3, San Francisco, March 1996, p.1242-1250
[NIE90] G. Niestegge, *"The Leaky Bucket Policing Method in ATM Networks"*, Int. Journal of Digital and Analog Communication Systems. Vol. 3, pp. 187-197, 1990.
[RAG91] E.P. Rathgeb, *"Modeling and Performance Comparison of Policing Mechanisms for ATM Networks"*, IEEE JSAC, Vol. 9, No. 3, pp. 325-334, 1991.
[RF95] A. Romanov and S. Floyd, *"Dynamics of TCP traffic over ATM networks"*, IEEE J. Selected Areas in Comm., 13 (4), 1995, p.633-641
[ROB93] J. Roberts (Ed), *"Performance evaluation and design of multiservice networks"*, COST 224, Commission of the European Communities, October 1992, Final Report
[ROJ94] J.W. Roberts, *"Weighted Fair Queueing as a Solution to Traffic Control Problems"*, COST 242 MID-TERM Seminar, L'Aquila, Italy, Sep. 1994.
[ROL94] L. Roberts, "Enhanced Proportional Rate Control Algorithm (EPRCA)", ATM Forum contribution nº 94-0735R1, August 1994.
[RV91] J. Roberts and J.Virtamo, *"The superposition of periodic cell arrival streams in an ATM multiplexer"*, IEEE Trans. Comm., 39(2), February 1991, p.298-303,
[SKL94] A. Skliros, *"Characterizing the Worst Traffic Profile passing through an ATM-UNI "*, Proceedings of the 2nd IFIP Conference on Performance Modelling and Evaluation of ATM Networks, Bradford (U.K.), 1994.
[SUM88] S. Sumita, T. Ozawa, *"Achievability of Performance Objectives in ATM switching Nodes "*, Int. Seminar on Performance of Distributed and Parallel Systems, pp. 45-46, Kyoto, Japan, Dec. 1988.
[TUR86] J. Turner, *"New Directions in Communications (or which way in the information age?)"*, Zurich Seminar on Digital Communications, pp. 25-32, March 1986.
[TUR96] J.S. Turner, *"Maintaining high throughput during overload in ATM switches"*, Proceedings of IEEE INFOCOM '96, Vol.1, San Francisco, March 1996, p.287-295
[UNI3.1] ATM Forum, ATM User-Network Interface Specification, September 1993.
[VR89] J.T.Virtamo and J.W. Roberts, *"Evaluating buffer requirements in an ATM multiplexer"*, Proceedings IEEE Globecom 89, 1989

[WAL91] E. Wallmeier, T. Worster, *"A Cell Spacing and Policing Device for Multiple Virtual Connections on one ATM Pipe"*, Proc. RACE R1022 Workshop on ATM Network Planning and Evolution, London, 1991.
[VER91] D. Verma, H. Zhang and D. Ferrari, *"Guaranteeing Delay Jitter Bounds in Packet Switching Networks"*,Proc. Tricomm´91, Chapel Hill, NC, pp. 35-46, April 1991
[YIN94] N. Yin and M. G. Hluchyj, *"On Closed-Loop Rate Control for ATM Cell Relay Networks"*, IEEE Infocom´94, pp.99-108.
[ZHA91] L. Zhang, *"Virtual Clock: A New Traffic Control Algorithm for Packet Switching Networks"*, ACM Transactions on Computer Systems, Vol. 9, No. 2, pp. 101-124, 1991.