Report Number: WUCS-94-13

1994-01-01

Congestion Control in ATM Networks

Apostolos Dailianas and Andreas Bovopoulos

Congestion Control in ATM Networks

Apostolos Dailianas and Andreas Bovopoulos

WUCS-94-13

May 1994

Department of Computer Science
Washington University
Campus Box 1045
One Brookings Drive
St. Louis MO 63130-4899

This work was supported by the National Science Foundation under grant NCR-9110183 and an industrial consortium of Ascom Timeplex, Bellcore, BNR, DEC, Gold Star, Italtel SIT, NEC America, NTT and SynOptics Communications.

Congestion Control in ATM Networks *

Apostolos Dailianas[†]
Washington University

Andreas Bovopoulos[‡]
Chipcom Corporation

1. Introduction

The development of new applications requiring high bandwidth coupled with technological advances in switching and transmission technology has promoted the idea of Broadband Integrated Services Digital Networks (B-ISDN). Asynchronous Transfer Mode (ATM) has been selected by ANSI and CCITT as the underlying transport technology for B-ISDN. ATM networks are packet-oriented, connection-oriented, sequence-preserving networks that can seamlessly transport many different media types (voice, video, data etc.). One of the main reasons for selecting ATM is the ability to provide access to an extremely wide range of applications. This access is achieved through the flexibility of supporting services with a variety of bandwidth demands and quality of service requirements. An equally important reason for selecting ATM is the ability to provide support for future services with characteristics which are not yet known, along with the inherent scalability of the total throughput of the ATM network.

An important feature of ATM is the bandwidth-saving capability that can be obtained by exploiting the statistical characteristics of the cell streams transmitted through the network. On the other hand, this is exactly the cause of network resource congestion. The contention for network resources introduces the need for mechanisms to prevent and to recover from congestion; otherwise severe degradation of the quality of service *(QoS)* can occur. This degradation is expressed as an increase in both the delay and the loss experienced by the cells passing through the network.

The mechanisms deployed against congestion can generally be distinguished as *reactive* and *preventive*. Reactive mechanisms are activated upon the appearance of congestion in the network, and their most common form is to limit the traffic entering the network until the network recovers from congestion. Reactive mechanisms have successfully been applied

[†]Computer & Communications Research Center, Bryan Hall 405 – Campus Box 1115, Washington University, One Brookings Drive, St. Louis MO 63130-4899, e-mail: apostolo@dworkin.wustl.edu, tel:314-9354163

[‡]Chipcom Corporation, 118 Turnpike Road, Southborough, MA 01772-1886, e-mail: abovopou@chipcom.com, tel:508-4905602

*This work was supported by the National Science Foundation under grant NCR-9110183, and an industrial consortium of of Ascom Timeplex, Bellcore, BNR, DEC, Goldstar, Italtel SIT, NEC America, NTT, SynOptics Communications.

in low-speed networks but they seem inappropriate for high-speed networks. This is due to the fact that in low speed networks, queueing delays are much larger than propagation delays. As a result, the sources are able to react before the state of the system changes significantly. In high-speed networks the propagation delay dominates, and the time taken for congestion to occur is on the order of queueing delays. As the state of the individual nodes can change rapidly, feedback information, which arrives at the source in twice the end-to-end propagation delay time after the time at which the cell that experienced congestion was sent, is outdated by the time it reaches the source, since the time to propagate the information is much larger than the queueing delays. As thousands of cells can be in transit across a link due to the high transmission speeds, the reactive mechanisms are effective only if the node buffers are unrealistically large, or the link utilization is very low. Clearly, end-to-end reactive control does not seem suitable for ATM networks. On the other hand, reactive control based on feedback between adjacent nodes that are not separated by large propagation delays can be effective, even in high speed networks [21, 37].

Preventive mechanisms try to predict the effects of congestion, for example through the prediction of the loss probability encountered by a stream, and take the necessary actions to prevent it. Examples of such preventive actions are intelligent route selection, rejection of a new call and control of the traffic entering the network. Preventive control mechanisms usually consist of two parts: the *admission control mechanism* which examines the current state of the network and accepts a new connection only if the new state after the source has been accepted will not violate the service contracts with all the sources (including the new source), and the *policing mechanism* which ensures that the admitted sources conform to a cell generation pattern agreed upon by a contract between the user and the network at the time when the source is accepted in the network. Both of these mechanisms have received much attention in recent years, but no completely acceptable mechanism has yet been developed.

This chapter mainly concentrates on issues related to predictive congestion control mechanisms. The rest of the chapter is structured as follows. Section 2 defines the quality of service (QoS) and gives a brief classification of applications according to their QoS requirements. Section 3 examines the problem of modeling the traffic characteristics of the sources entering the network. Section 4 examines the different resource allocation mechanisms that can be provided by the network and provides an event-driven specification of the actions performed by such mechanisms. In Section 5 the admission control task is presented, in close relation with the mechanisms provided by the network. Section 6 examines the distortion of the original traffic pattern as it traverses the network. Section 7 presents different approaches to ensuring the legal operation of the sources through the application of policing mechanisms.

2. Quality Of Service (QoS) - Classification of Services

The ATM network is shared by media streams with very diverse traffic characteristics and performance requirements. A QoS specification which is common to all existing services consists of the specification of the loss and delay requirements of the service.

The loss requirements are often dictated by the information content of the sequence of cells transmitted by the source. Some virtual circuits (VCs) transmit correlated sequences of cells, called *bursts,* generated because of the physical (source dependent) or logical (protocol dependent) structure of the information stream carried by the VC. Losing a number of cells of a burst may result in the loss of the information content of the whole burst. Therefore, an ATM network should support VCs requesting QoS guarantees with respect to the burst-loss probability [7, 33, 50].

To support applications such as voice or video transmission, an ATM network should also efficiently support VCs requiring performance guarantees at the cell level. Providing guarantees for the loss at the cell level is the most studied QoS aspect appearing in literature. Finally, to support applications such as file transfer or distributed databases, the ATM network should support VCs demanding a loss-free connection.

The specification of QoS with respect to loss usually involves extremely small numbers. A typical range of requested cell or burst loss probability extends from 10^{-2} to 10^{-9}. Having to deal with such numbers suggests that the loss parameters involved in the specification of QoS should be defined as the statistical measure of the performance of the aggregate traffic of all the sources requiring the same QoS.

The second type of guarantees that the network should provide concerns the delay that the cells experience when passing through the network. The most common approach appearing in literature is to guarantee an upper bound for the delay that a cell will experience. In a multi-media applications environment, such an approach is not sufficient, since continuous media *(CM)* streams, such as real-time video, have stringent delay jitter requirements. The term *jitter* refers to the distortion introduced to the original traffic pattern by the consecutive queueing stages along the path of the connection. It is quantitatively measured as the maximum difference between the end-to-end delays of any two cells of the same stream. For such VCs, an ATM network must provide jitter guarantees, independent of the number of nodes the media stream traverses from source to destination. Examples of network designs that provide delay guarantees according to the delay requirements of each user can be found in [4, 27], while designs that provide strict jitter guarantees and a constant end-to-end delay can be found in [15, 23, 24, 48].

A brief classification of services according to their QoS requirements can be found in [25].

3. Traffic Modeling

To develop an analytical method or to obtain simulation results concerning congestion in a network, a mathematical description of the traffic entering the network must be developed. Since the streams entering the network can have widely different traffic generation patterns, it seems appropriate to develop a different *source model* for every group of sources with similar characteristics. Such a source model should be able to capture all the essential characteristics of the source that are related to the way it produces cells. For example, a two-state Markov chain (on-off source) is commonly used to model bursty sources. Traffic,

like video or voice, which is correlated, can be modeled through a first-order autoregressive process [32, 37, 44] or an interrupted Poisson process [30]. Variable-bit-rate sources such as compressed video can be modeled using a discrete, finite state, continuous time Markov process. It is also very useful to develop models for the traffic produced by the superposition of sources. For example, the superposition of voice and data sources can be characterized through the index of dispersion for intervals *(IDI)* [46] and can also be modeled by a Markov modulated Poisson process *(MMPP)* [28], while the superposition of on-off sources and continuous-bit-rate sources can be modeled as a stepwise variable-bit-rate source.

The description of a source through a set of traffic parameters is the basis for resource allocation, admission control, pricing and shaping/policing. All of these operations must be performed using a very limited set of parameters, due both to computational constraints (for admission control) and hardware implementation constraints (resource allocation, policing/shaping). This introduces the need to describe a source through a limited set of parameters applicable to all sources, referred to as the *traffic descriptor.* An inherent problem of a common description for all sources is that some of the information concerning the statistical characteristics of each source is lost, since many different sources are mapped to the same traffic-descriptor parameters. Such a loss of information can lead to inefficient resource allocation and consequently decreased network utilization.

3.1. Traffic Descriptor

To constitute a traffic descriptor, a set of traffic characteristics should process as many as possible of the following properties:

1. It should accurately capture the statistical behavior of diverse sources.
2. It should correspond as much as possible to parameters whose impact on loss, congestion, and buffer dimensioning is quite clear.
3. The parameters should be easy to measure and control (observability and controllability). They should also be easily understood and specified by the user.
4. It should impose the fewest constraints on the traffic stream. It is thus desirable to characterize the traffic streams parametrically rather than classifying them into a predefined set of traffic classes.

The most common set of parameters used to describe a source consists of the *mean rate*, the *peak rate* and the *mean burst length.* The mean rate is sometimes replaced by the *burstiness* defined as the ratio of the peak over the mean rate. Examples of the description of typical applications such as telephony, file transfer, video and others, through this set of parameters can be found in [22]. Frequently, only a subset of the above set, namely the peak rate and the mean rate, is used in the literature.

The main problem with the above set of parameters is that it does not satisfy the requirement for observability and controllability. This is a problem common among all

specifications whose set of parameters includes average values of variables with no specification of the averaging interval. Recently, a new complete version of the above set of parameters to characterize a source has been proposed, focusing on the need for the observability and controllability of the parameters constituting the traffic descriptor. It consists of four parameters, namely the *peak rate*, the *cell delay variation tolerance*, the *maximum sustainable rate* and the *maximum compliant burst at rate R*. The maximum sustainable rate is intended to describe the variable behavior of a traffic source in order to allow statistical multiplexing. The maximum compliant burst at rate R essentially replaces the mean burst length, which is not easy to observe or control.

Another set of parameters which is sometimes used, consists of the *minimum interarrival time*, the *average interarrival time*, and the *averaging interval* [52]. Although more limited than the previous set of parameters, they are easy to observe and control due to the specification of the averaging interval over which the parameters are specified.

4. Resource Allocation - Runtime Specification

The same network resources (buffers and bandwidth) are shared among all traffic streams contending for them. To design the hardware mechanisms that take the appropriate actions in case of congestion and also for the purposes of the connection admission algorithm, an event-driven specification of the actions to be taken at each node in case of congestion is needed. This specification is called the *run-time specification* of the system.

The original idea in ATM networks was to share the network resources among the users through statistical multiplexing, taking advantage of the diversity of the traffic characteristics of the multiplexed sources. Statistical multiplexing seemed to be the most promising of the bandwidth allocation techniques with respect to bandwidth utilization due to the exploitation of the burstiness of the streams passing through the network.

In the case of statistical multiplexing, the run-time specification is very simple and does not require any special hardware support. Whenever a cell arrives, it enters the buffer if the buffer is not full; otherwise it is dropped. Statistical multiplexing has received great attention in the literature [5, 19, 22], and some associated drawbacks have been revealed. The two most important problems associated with statistical multiplexing are that (i) excessive computations are needed to compute the QoS provided, thus preventing real-time decision for the acceptance or the rejection of a new call, and (ii) a different QoS cannot be provided to sources with different demands.

To overcome the first problem, namely the excessive computations required, to enable the evaluation of the provided QoS in real-time, some upper-bounding approximations have been proposed. They are presented in Section 5.

The other important problem of statistical multiplexing is that it fails to guarantee the different QoS requirements that different applications have. For example, it cannot provide different cell-loss probability to two otherwise identical sources with different loss requirements. Furthermore it cannot provide any guarantees on the burst loss probability or on the delay aspects of the QoS specification.

Other allocation schemes, which typically focus on a subset of the QoS parameters, have been proposed in literature. They include *peak rate allocation, minimum throughput allocation* [51], *stop-and-go* queueing discipline [23, 24], *fast buffer reservation (FBR)* [50], *fast reservation protocol (FRP)* [7] and others. The first three focus on guaranteeing the QoS requirements at the cell level, while the last two focus on the integrity of the bursts.

In peak rate allocation, the maximum rate for every source is allocated at every switch. No specific hardware support is needed to perform the allocation. The source is accepted at a node only if the cumulative rate of all the accepted sources is less than the link rate. Peak rate allocation is very easy to implement, provides loss-free transmission to conforming users, but leads to low resource utilization in the presence of bursty traffic.

In minimum throughput allocation, the network guarantees to the user a minimum throughput, even in case of congestion; any excess rate is not guaranteed. A simple runtime specification is based on the idea that in case of congestion, a throughput which is a percentage β of the link rate can be guaranteed by allocating the same percentage β of buffer slots to the virtual circuit *(VC)*. Upon arrival of a cell a counter associated with the VC is increased by one; upon departure, it is decreased by one. In case of congestion, an arriving cell is accepted only if the counter does not exceed the number of buffer slots allocated to the VC. A more sophisticated mechanism that guarantees each flow its reserved average throughput is called *Virtual Clock* [51].

The stop-and-go queueing discipline [23, 24] is a peak rate allocation approach focusing on the provision of strict delay jitter guarantees. Time is divided into frames, and every VC is assigned a number of cells it can transmit per frame. Upon the arrival of a cell, if the specified number of cells per frame is not exceeded, the cell is accepted and transmitted at the next eligible frame of the outgoing link; otherwise it is dropped. The drawbacks of this scheme are the low utilization, in the case of a bursty source, and the increased complexity introduced by framing.

Both the *fast buffer reservation* scheme *(FBR)* and the *fast reservation protocol (FRP)* focus on the integrity of a burst and are based on the assumption that during each burst, the source requires a constant bit rate. The required buffer/bandwidth is allocated to the connection for the whole burst duration and released after the completion of the burst. In both schemes, only the burst that would cause congestion to arise is dropped, while the rest of the bursts are not affected.

The FBR scheme [10, 50] associates with each virtual circuit a share of the cell slots that is equal to the ratio of the peak rate of the source to the link rate. Acceptance or rejection of cells is performed on a burst basis. If upon the arrival of a burst the number of buffer slots associated with the source is available, then the burst is accepted, and the associated number of slots is reserved for the whole duration of the burst; otherwise the whole burst is dropped. The allocation of the buffer proportionally to the peak rate of the source implicitly allocates the link bandwidth in exactly the same way. In case of congestion, the system behaves like a peak rate allocation scheme: none of the admitted bursts experiences any loss. The loss probability is moved from the probability that a cell is lost due to buffer overflow to the probability that a burst is rejected access to the buffer. The drawbacks of

FBR are the need to recognize the beginning and end of a burst and the lower bandwidth utilization than statistical multiplexing for traffic streams composed of small bursts.

The fast reservation protocol [7] takes into account the fact that the statistical characteristics of the traffic streams that are generated by data transfer calls and LAN-LAN interconnections are not known in advance. The sources are modeled as stepwise, variable-bit-rate sources. Every stepwise increase in the rate can be considered as the the beginning of a new burst. A source must wait before increasing its activity, until the corresponding stepwise increase of bandwidth has been reserved in all switching elements along the connection. While the negotiation is in progress, the source may still transmit at the previous rate. This scheme, called *FRP/DT* (DT: Delayed Transmission), introduces transmission inefficiency, since the reserved bit rate cannot be used before a round-trip delay. An alternative to FRP/DT is *FRP/IT* (IT: immediate transmission), which is intended for applications which cannot wait for network agreement before changing their traffic activity. The burst is sent along the connection without previous agreement. The bit rate is negotiated on-the-fly at each switching element; if rejected, the whole burst is discarded. Both FRP/DT and FRP/IT do not provide any guarantee that once a burst of a connection has been accepted, subsequent bursts will also be accepted.

There are two main concerns with FBR and FRP. The first is that there are applications like video and voice where it is desirable that the loss of cells be as distributed as possible. The second is that in order to preserve the burst integrity, some of the bursts are rejected, leading to reduced link utilization since part of the bandwidth is not utilized.

5. Admission Control

Admission control refers to the decision whether to accept an incoming call, based on the traffic characteristics and the QoS requirements of all the already accepted sources, as well as traffic characteristics and the QoS requirements of the new call. The admission control algorithm is closely coupled with the way the underlying network reacts to congestion, as described in Section 4, in the sense that the decision algorithm is dictated by the resource allocation policy that is chosen. The objective of an admission control mechanism is to guarantee that there are sufficient resources to realize the contracts with each of the admitted sources, under the assumption that the new call is accepted.

To decide whether to accept a new connection, it is necessary to determine for all nodes, from the source to destination, if the acceptance of a source will either fail to guarantee the required QoS for the new source or will lead to a QoS violation of the already admitted sources. In either case, the incoming call must be rejected. The problem is decomposed into applying the decision algorithms at each intermediate node. Ideally, a Connection Admission Control *(CAC)* algorithm should be able to accurately predict violations of the requested QoS, while coping with priorities and multimedia traffic and also operating in real time. The use of traffic descriptors that can easily be policed is also important, to prevent violating users from affecting the service provided to the rest of the users. Policing of traffic is discussed in Section 7.

One way to classify CAC algorithms is by means of the criteria used for acceptance/rejection of an incoming call. The two most commonly used approaches are the *QoS evaluation method* and the *equivalent bandwidth method.* In the QoS evaluation method, the acceptance criterion is the QoS which can be provided to the source under a given set of resources (buffer and bandwidth) and a given set of active sources. The QoS is evaluated for all the sources including the new source. If acceptance of the new source leads to a violation of the contract with any of the already accepted sources or failure to guarantee the requested quality for the new source, then the new source is rejected. Examples of application of the QoS evaluation method abound [4, 15, 16, 19, 42, 41, 50].

In the *equivalent bandwidth method* the acceptance criterion is the existence of link bandwidth equal to the equivalent bandwidth. The *equivalent bandwidth* is defined as the bandwidth required to satisfy the requested QoS, given a particular buffer size and set of active sources. The use of equivalent bandwidth is exemplified in [1, 17, 22]. In [17, 22] simulation is used to obtain the equivalent bandwidth for the case of homogeneous on-off sources with a given mean rate, peak rate and mean burst length. The equivalent bandwidth method is effective and simple for the case of homogeneous sources. For heterogeneous sources, however, it becomes very difficult to determine the precise bandwidth needed for each source. Approximate methods must be used to estimate bandwidth requirements. Many of the approximate methods try to use the analysis for the homogeneous case to approximate the non-homogeneous case. One of the simplest, called the *linear approximation method* [1], assumes that the bandwidth needed for a source is the same as the bandwidth that the source would require if the whole available bandwidth were occupied by homogeneous sources of the same type as the source under consideration. It has been shown by simulation that the linear approximation method underestimates the bandwidth needed by a source. A similar approach appears in [20]. The homogeneous analysis is applied under the assumption that all sources have the same characteristics as the most bursty of the sources of the traffic mix.

Table 1: Classification of CAC algorithms

	algorithm input	algorithm output
QoS evaluation method	Network resources Established calls New call	QoS of all calls
Equivalent bandwidth method	Established calls New call	Bandwidth required to support new call

The major advantage of the equivalent bandwidth approach is the determination of the amount of extra bandwidth needed to accommodate the new source. This can be useful, for example, in the case of a virtual path, where, although there may not be enough bandwidth to support the incoming call, the exact amount of required bandwidth can be determined. Equivalent bandwidth could also be used directly for pricing purposes. On the other hand,

the QoS evaluation method provides an estimate of the QoS the source will receive. This estimate can be used as a negotiation parameter between the user and the network.

Both the QoS evaluation method and the equivalent bandwidth method are computationally very expensive and preclude real-time implementations of the admission algorithms. Several techniques can be used to reduce the computational complexity. One such technique recomputes the QoS characteristics provided to the sources under a specific traffic mix, namely a traffic mix chosen to be representative of the applications supported by the system. A region called the *admissible load region* is then specified. Its boundaries correspond to the maximum number of sources that can be admitted without violating the requested QoS. The admission decision is based on the admissible load region. Examples of this technique can be found in [29].

One technique that tries to simplify computations by introducing approximations to the traffic pattern is called the *fluid-flow approximation* technique [18, 47]. The traffic pattern is approximated with a Markov modulated fluid flow, in which the rate of fluid generation is determined by the state of the controlling continuous-time Markov chain.

A common technique used by most CAC algorithms to simplify computations is to relax the assumption that the rate of the source can be arbitrary. The rate of the source can only be a *multiple of a fundamental rate*. In [31] it is claimed that a quantization of the rates would have only a small impact on the decrease of throughput and the increase of the blocking probability. This technique has extensively been used in literature [4, 15, 16, 50].

Another common technique involves the use of *recursive equations* to calculate the new values of quantities involved in CAC when a new call requests connection or when a call is disconnected. Examples of recursive algorithms can be found in [4, 10, 15, 16, 19, 50].

Finally, in terms of techniques to reduce the computational complexity, an approach which assumes that the multiplexors are *bufferless* is proposed in [19]. This assumption dramatically reduces the amount of computations that needs to be performed, thus making real-time implementations possible. The computations give an upper bound of the loss probability that a source will experience (because of the bufferless assumption). The connection admission decision is safely based on this upper bound. Examples of this approximation can be found in [4, 15, 19].

In Section 4 some examples of the resource allocation schemes used by the underlying network are presented. The CAC algorithm must take the resource allocation scheme into account since it determines what happens in case of congestion. Algorithms assuming that the underlying network provides pure statistical multiplexing can be found extensively in literature. Examples include [1, 2, 4, 5, 15, 16, 19, 22]. In [43, 42, 41] the tendency of users to overestimate their traffic characteristics is recognized, and simple traffic estimation performed by the network is used instead of the user-provided description. Examples of algorithms assuming priorities at the cell level can be found in [4, 15, 41]. In [15, 16, 48] algorithms for a network using the stop-and-go queueing discipline are presented. The algorithm for fast buffer reservation is presented in [50]. Notice that no algorithm is presented for the scheme operating under the fast reservation protocol, since in this case the admission is performed on a burst basis and no performance guarantees are provided on a call basis.

CAC algorithms heavily depend on the traffic characteristics of the source. Since the decisions are valid only for the negotiated traffic characteristics, it should be ensured that they are respected at all times. Violation of the declared traffic characteristics might be due to two factors. The first is the change in the traffic characteristics introduced by the successive queueing stages within the network; this is discussed in Section 6. The second is that the traffic produced by the user is by itself in violation of the contract. In this case, appropriate mechanisms discussed in Section 7 are used to restrict violating users.

6. Change of Traffic Characteristics

When a traffic stream traverses the network, its structure can undergo significant alterations [8, 23, 52]. This is due to the random delays a cell experiences while progressing along a path. These random delays are caused by queueing in each multiplexing stage. The phenomenon of the alteration of the initial time structure of the cell stream is called *Cell Delay Variation* (*CDV*). For example if the FCFS service discipline is used at each node, two cells with small inter-cell arrival times can exit spaced out or close together, depending on the total workload that has arrived in the interim on the other connections sharing the same output link. The situation in which two cells come close to each other is described by the term *clumping* or *clustering,* while the opposite effect is described by the term *dispersion.*

From the point of view of the users, especially in a multimedia environment involving continuous media streams such as real-time video, the distortion of the traffic characteristics is undesirable, since it leads to the violation of the stringent jitter requirements of such streams. Also, from the point of view of the network, the clustering effect can lead to an increase in the peak rate of the streams at downstream nodes of the network. This introduces a need for description of the traffic pattern distortion along the path of a connection; otherwise the results obtained by the CAC algorithm might not hold. The clustering effect also implies that the amount of resources allocated to each stream has to increase at downstream nodes.

Several approaches trying to characterize the traffic pattern distortion along the path of a connection appear in the literature [12, 13, 35]. One of the problem of these solutions is that the set of equations that need to be solved may become unsolvable when traffic forms traffic loops. Moreover, most of them refer to networks with work-conserving service disciplines, *i.e.* networks where the link cannot be idle when there are packets waiting in the queue.

There are several solutions regarding the prevention of the alteration of traffic characteristics. In [8] it is suggested that the initial structure of the source is reconstructed at every node through the use of dedicated hardware mechanisms. The proposed hardware mechanism is called the *cell spacer-controller* and can also be used for policing purposes. Its operation is described in Section 7, which refers to policing.

In [23] it is suggested that for the purposes of providing delay jitter guarantees to the streams, it is sufficient to restrict the distortion of traffic characteristics through the introduction of a framing strategy called *stop-and-go queueing* . In the proposed framing

strategy the time is divided locally (at the output links of each node) into equal time intervals called *frames*. The traffic characterization in this scheme consists of specifying the maximum number of cells that can arrive within each frame. With reasonable hardware complexity, the proposed scheme guarantees that any two cells contained at the same frame at the source will also be contained at the same frame at the destination, independent of the number of nodes the traffic goes through, but at the cost of increased end-to-end delay. Moreover, the proposed scheme guarantees that the traffic characterization which is valid at the source will also be valid throughout the network. This property makes the proposed scheme particularly suitable for the purposes of CAC algorithms. An extension of the above scheme, aimed at the provision of different restrictions on the extent of the permissible traffic distortion, according to the distortion the source can tolerate, is proposed in [24]. The extension is based on the use of frames with different sizes at the expense of increased hardware complexity.

7. Usage Parameter Control (Policing)

Given that resource allocation is performed through negotiation between the user and the network, based on the source description using a traffic descriptor, the parameters of the descriptor need to be policed, so that any change in the user's behavior does not affect the network performance as seen by other users. The parameters that need to be policed are those involved in the CAC algorithm. Ideally, a policer should be completely transparent to conforming traffic in terms of all QoS characteristics that are of interest to the user. The policer should also be able to immediately detect violating traffic. The design of a policing mechanism can be broken into three tasks: (i) the design of an algorithm for the *detection of violating traffic,* (ii) the policy for the *handling of violating traffic,* and (iii) the decision as to where to *place the policer.*

In terms of designing algorithms to *detect violating traffic,* the attention in the literature has been given to the peak rate and the mean rate, since they are the most commonly used parameters by CAC algorithms. Another parameter whose impact on the network performance has been shown to be important through simulations is the mean burst length. While quite a few policing mechanisms have been suggested for policing of the peak rate and the mean rate [8, 49], most policers can impose limits only on the maximum burst length but are unable to police the mean burst length. A desirable property of a well-designed policer is that it is *rule based, i.e.* it bases its decisions on algorithms that are deterministic and known to the user in advance, so that the user knows at any time with certainty whether the user is in violation and can take appropriate action.

With regard to policing parameters like the mean rate of a source, a problem arises because of the averaging nature of the parameters to be policed, namely that the ideal measuring interval for the particular parameter should be infinite. Obviously this cannot be the case since detection of violating traffic would require infinite time. One solution to this problem is the introduction of a *finite measurement interval* or an *averaging interval* over which the parameter is monitored. A typical example of a finite measurement interval can be found in [40]. One of the main difficulties with such an approach is determining the

length of the measurement interval, which influences both the amount of buffer needed by the policer and also the time required to detect violating traffic. Long intervals need big buffers and lead to slow response times in case of violation, but they give a more accurate estimation of the mean. Long intervals might be required, for example, in the case of bursty sources with long mean burst lengths, where the measuring interval for the mean rate must be a multiple of the mean time between successive bursts. On the other hand short intervals can use small buffers and provide short response times, but they give inaccurate estimation of the mean, thus increasing the probability of blocking conforming traffic.

In terms of *handling violating traffic,* two approaches can be used. The first is to *discard* violating traffic, making sure that it will not interfere with conforming traffic at any stage in the network. The second is to *tag* violating traffic, indicating that tagged cells should be transported but can be discarded if they encounter congestion along the connection path. Tagging of cells can be performed by setting the priority bit of the cell header to low priority. Notice that the tagging approach cannot be used in a framework supporting priorities, since tagged traffic would interfere with conforming low priority traffic.

Finally, the decision of *where to place the policer* will affect the traffic expected at the entrance to the policer. CCITT's recommendation I.371 [9] suggests that connections be monitored and controlled in terms of offered load, by a mechanism called the usage parameter control *(UPC)* mechanism at the user-network interface *(UNI)* at the T_B reference point, and by a mechanism called the network parameter control *(NPC)* at the network-node interface *(NNI)* point. The main reason for the distinction between the two mechanisms is that while the traffic characteristics at the T_B have suffered only the distortion due to any device between the source and the UNI, the distortion at the NNI point is much bigger due to the effect of the cell delay variation (CDV) (see Section 6) introduced by upstream networks. This recommendation points out that a certain CDV tolerance between the source and the T_B reference point must be specified for the UPC mechanism properly to control the parameters agreed on during negotiation time. The results presented in [12, 13] suggest that the CDV tolerance at the T_B access point can be upper-bounded by the CDV introduced by a single-stage multiplexer, while the results presented in [38], considering realistic bursty sources, indicate that such an assumption is not always valid. The placement of the policer suggested by CCITT, although the most natural choice, does not solve the problems arising when the cell delay variation introduced at every node makes the traffic violate the agreed upon characteristics at intermediate nodes.

Two examples of policing mechanisms are the *leaky-bucket* and the *cell spacer-controller.* Their target policing parameters are the mean rate and the peak rate of a connection respectively. The *leaky bucket* [49] is a counter associated with a user. Every time the user transmits a cell, either the counter is incremented and the cell is transmitted to the network, or, if the counter has reached a threshold value, the cell is discarded and the counter value remains unchanged. At periodic intervals of time the counter is decremented to zero. The rate at which the counter is decremented determines the mean rate and is specified by the user. The value of the threshold is also specified by the user and is intended to control the user traffic burstiness. The leaky-bucket enforces the mean rate and poses a limit on the maximum burst length. Some problems associated with the leaky bucket along with proposed solutions, can be found in [21].

The leaky-bucket and other mechanisms such as window flow control mechanisms are examples of a general category of flow control mechanisms called *pick-up mechanisms.* Pick-up mechanisms consist of a pool of N credits generated according to a rule depending on the mechanism. Each incoming cell takes a credit if available; if not, it is detected as a violating cell. Pick-up mechanisms do not reshape the conforming traffic; thus they do not introduce any extra latency. An extension to the leaky bucket which introduces possible shaping of the traffic was proposed in [45]. The proposed mechanism, called *buffered leaky bucket,* introduces a buffer whose purpose is to store cells if no tokens are currently available. Pick-up mechanisms are very attractive because of their simplicity.

Notice that due to CDV at both the T_B and the NNI reference points, special care has to be taken in pick-up mechanisms not to reject conforming traffic. Pick-up mechanisms may have to contain a few tens of credits in order to be transparent to conforming traffic, while accommodating for CDV. According to a heuristic rule given in [34], the leaky bucket policing a connection with a peak emission period T must contain $[d/T]+2$ credits, in order for the rejection probability of conforming cells to be very low (say 10^{-10}). The parameter d is called the 10^{-10} quantile of the random delay of the stream, and its value is such that for the random delay W experienced by a cell, it holds that $Pr\{W > d\} < 10^{-10}$. Pick-up mechanisms must observe a few tens of cells before taking any action, introducing some latency in the mechanism's reaction.

A mechanism intended for policing the peak rate is presented in [8, 26]. The hardware implementation suggested in [8] is called *spacer-controller.* The basic idea of this mechanism is to store the incoming traffic and re-transmit it in such a way that the traffic pattern is as close as possible to the initial structure of the source in terms of the distance between consecutive cells. At the same time, violating cells are discarded. The decision for discarding a cell is based on how close in time it arrives with respect to the previous cell. If the time difference from the previous cell plus a time τ (related to the CDV) is smaller then the minimum expected spacing between consecutive cells, the cell is discarded since most probably it was in violation in the first place; otherwise the cell is scheduled to be transmitted with the appropriate spacing from the previous cell of the connection. The main design parameter in this scheme is the time constant τ which corresponds to the jitter associated with the CDV of the connection. Heuristic arguments are used to dimension τ in order for the probability to discard conforming cells to be negligible.

The spacer-controller is an example of a general category of mechanisms called *shapers.* Shaping a connection consists in buffering cells of a connection and then re-transmitting the cells so that the traffic of the connection is as close as possible to that generated by the source. In shaping mechanisms, cells are systematically buffered. A shaper may modify conforming as well as non-conforming traffic.

Notice that apart from being a policing mechanism, the space controller can also be used as a shaping mechanism that compensates for the distortion introduced by the CDV effects. The clumping effect (cells coming close to each other) is not totally eliminated but is drastically limited. On the contrary the dispersion effect is increased, but this is a minor problem however since, in contrast with clumping, dispersion does not have a drastic impact on the performance of the network. The additional dispersion introduced by the cell spacer is similar to the dispersion introduced by one multiplexing stage.

Apart from restoring the initial traffic pattern, shaping of the traffic of a source can also be used to make the traffic smoother and more appropriate for statistical multiplexing, resource allocation and traffic description. In this case the shaper should guarantee zero loss. To achieve the goal of smoothing the traffic, the shaper has to introduce some delay. It can thus be very effective for sources that do not have stringent delay requirements. Finally, another use of a shaper can be to shape the traffic of the source at the customer premises, before the policer, so as to make the source traffic conform to the predefined traffic parameters [40].

Policing mechanisms can be combined in parallel or in tandem when multiple parameters must be simultaneously policed [36]. For example, policing the peak and the mean rate can be performed by a spacer-controller followed by a leaky-bucket. Also if a stream contains cells with different priorities, two (or more) policing mechanisms can be combined in parallel, one for each sub-stream. In this case, special care must be taken to maintain the initial ordering of cells.

Finally, to compare different policing mechanisms, a simple measure of the effectiveness of a policing function is needed. It has to take into account the percentage of cells that were discarded although they could have been transmitted, the probability that a legally operating source is restricted, the time it takes to respond to a violating source, and other parameters. Such measures of effectiveness are given in [3, 6, 39]. According to the definition of efficiency proposed in [39], the leaky bucket turns out to be the most efficient policing algorithm.

8. Summary

ATM will support a variety of services with widely different traffic characteristics and QoS requirements. Efficient use of the network introduces the need for the sharing of resources in a way which can lead to congestion and the deterioration of the QoS seen by the users. This chapter reviews congestion control and avoidance methods, focusing on preventive control and especially on resource allocation schemes, admission control algorithms and policing of the behavior of sources.

References

[1] S. Akhtar "Congestion Control in a Fast Packet Switching Network", Master Thesis, Washington University in St. Louis, Dec. 1987.

[2] A. Baiocchi, N. Blefari, M. Listanti, A. Roveri, R. Winkler "Modeling Issues on an ATM Multiplexer within a Bursty Traffic Environment", Proceedings of IEEE INFOCOM'91, pp. 83-91.

[3] A. Berger, A. Eckberg, T. Hou, D. Lucantonni, "Performance Characterization of Traffic Monitoring, and Associated Control Mechanisms for Broadband Packet Networks", Proceedings of the IEEE GLOBECOM, Dec. 1990, Vol. 1, pp. 350-354.

[4] G. Bianchi, V. Trecordi "Proposal for a Comprehensive Bandwidth Management Scheme and Connection Acceptance Rule for B-ISDN", Technical Report WUCS-92-24, Washington University in St. Louis.

[5] M. Bonnati, A. Bovopoulos, A. Dailianas, A. Gaivaronski "Exact and Approximate Solution of a Multiplexing Problem", Technical Report WUCS-92-09, Washington University in St. Louis.

[6] Borgonovo, L. Fratta "Policing in ATM Networks: an Alternative Approach", 7th ITC Seminar, Morristown, 1990.

[7] P. Boyer, D. Tranchier "A Reservation Principle with Applications to the ATM Traffic Control", Computer Networks and ISDN Systems Vol. 24, No 4, May 1992, pp. 321-334.

[8] P. Boyer, F. Guillemin, M. Servel, J. Coudreuse "Spacing Cells Protects and Enhances Utilization of ATM Network Links," IEEE Network Magazine, Vol. 6, No. 5, Sept. 1992, pp. 38-49.

[9] CCITT Recommendation I.371 "Traffic Control and Congestion Control in B-ISDN", Geneva 1992.

[10] J. Cox, J. Turner "Project Zeus: Design of a Broadband Network and its Application on a University Campus", Washington University in St. Louis, Technical Report WUCS 91-45.

[11] J. Cox, M. Gaddis and J. Turner "Project Zeus: Design of a Broadband Network and its Application on a University Campus", IEEE Network Magazine, Vol. 7, No 2, Mar. 1993, pp. 20-30.

[12] R. Cruz, "A Calculus for Network Delay, Part I: Network Elements in Isolation", IEEE Transactions on Information Theory, Vol. 37, No 1, Jan. 1991, pp. 114-131.

[13] R. Cruz "A Calculus for Network Delay, Part II: Network Analysis", IEEE Transactions on Information Theory, Vol. 37, No 1, Jan. 1991, pp. 132-141.

[14] R. Cruz, "A Calculus for Network Delay, Part I: Network Elements in Isolation", and "A Calculus for Network Delay, Part II: Network Analysis", IEEE Transactions on Information Theory, Vol. 37, No 1, Jan. 1991, pp. 114-131 and 132-141.

[15] A. Dailianas and A. Bovopoulos "Real-time Admission Control Algorithms with Delay and Loss Guarantees in ATM Networks," Technical Report WUCS-93-31, Washington University in St. Louis, and in proceedings of INFOCOM'94.

[16] A. Dailianas and A. Bovopoulos "Design of Real-Time Admission Control Algorithms with Priority Support," Technical Report WUCCRC-94-05, Washington University in St. Louis.

[17] M. Decina and T. Toniatti "On Bandwidth allocation to Bursty Virtual Connections in ATM Networks," IEEE Proceedings ICC'90, Atlanta, April 1990, Vol. 3, pp. 844-851.

[18] A. Elwalid, D. Mitra "Fluid Models for the Analysis and Design of Statistical Multiplexing with Loss Priorities on Multiple Classes of Bursty Traffic", Proceedings of INFOCOM'92, pp. 415-425.

[19] H. Esaki "Call Admission Control Method in ATM Networks", Proceedings of ICC 1992, pp. 1628-1633.

[20] J. Filipiak "Structured Systems Analysis Methodology for Design of an ATM Network Architecture," IEEE JSAC, Vol. 7, No 8, 1991, pp. 1263-1273.

[21] L. Fratta, L. Mucumeci, G. Gallasi, L. Verri, "Congestion Control Strategies in ATM Networks", European Transactions on Telecommunications, Vol. 3, No 2, March-April 1992, pp. 183-193.

[22] G. Gallassi, G. Rigolio, L. Fratta "ATM: Bandwidth Assignment and Bandwidth Enforcement policies", Proceedings of the IEEE GLOBECOM 1989, pp. 1788-1793.

[23] S. Golestani "Congestion-Free Transmission of Real-time Traffic in Packet Networks," Proceedings of INFOCOM'90, June 1990, pp. 527-536.

[24] S. Golestani "A Stop-and-Go Queueing Framework for Congestion Management," Proc. ACM Sigcomm'90, Vol. 20, No 4, Sept. 1990, pp. 8-18.

[25] Raman Gopalakrishnan, A. Bovopoulos "Design of a Multimedia Applications Development System," Technical Report, Washington University in St. Louis, WUCS-92-27.

[26] F. Guillemin, P. Boyer and L. Romoeuf "Spacer-Controller: Architecture and first assessments," Workshop on Broadband Communications, Estoril, Portugal, January 1992.

[27] S. Gupta, M. El Zarki "Traffic Classification for Round-Robin Scheduling Schemes in ATM Networks", Proceedings of the IEEE INFOCOM'93, pp. 820-827.

[28] H. Heffes and D. Lucantoni "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance", IEEE JSAC, Sept. 1986, pp. 856-868.

[29] J. Heyman, A. Lazar and G. Pacifici "Real Time Scheduling with Quality of Service Constraints," IEEE JSAC, Vol. 9, No 7, 1991, 1052-1063.

[30] I. Ide "Superposition of Interrupted Poisson Processes and its Application to Packetized Voice Multiplexors", Proceedings of the 12th ITC, Torino, Italy, 1988, Vol. 3, p. 3.1B.2.

[31] Chin-Tau Lea "What Should be the Goal of ATM?", IEEE Network Magazine, Vol. 6, No. 5, Sept. 1992, pp. 60-66.

[32] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, J. Robbins "Performance Models of Statistical Multiplexing in Packet Video Communications", IEEE Transactions on Communications, Vol. 36, No. 7, July 1988, pp. 834-844.

[33] S. Mahdavian and A. Bovopoulos "Effective Loss of Multiplexed ATM Cell Streams," Technical Report WUCS-93-08, Department of Computer Science, Washington University, and proceedings of INFOCOM'94.

[34] G. Niestegge "The Leaky Bucket Policing Method in ATM Networks", International Journal of Digital and Analog Communications Systems, Vol. 3, No 2, Apr-Jun 1990, pp. 187-197.

[35] Abhay Kumar J. Parekh, R. Gallager "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple Node Case", proceedings of INFOCOM'93, pp521-530.

[36] S. Proulx "Conception et Analyse de Fonctions de Surveillance de Source dans un Reseau de type ATM" by Stephane Proulx, May 1992, Dept. of EE, Ecole Polytechnique de Montreal.

[37] G. Ramamurthy, B. Sengupta "Predictive hop-by-hop Congestion Control Policy for High Speed Networks," Proceedings - IEEE Ifocom 1993, pp. 1033-1041.

[38] G. Ramamurthy, R. Dighe, "Performance Analysis of Circuit-Switched Traffic over a Packet Network", Proceeding of ICC, April 1990, pp. 572-578.

[39] E. Rathgeb "Policing Mechanisms for ATM Networks, Modeling and Performance Comparison," Proceedings of the 7^{th} ITC Seminar, Morristown, 1990.

[40] G. Rigolio, L. Verri, L. Fratta "Source Control and Shaping in ATM Networks", Proceedings of the IEEE GLOBECOM 1991, pp. 276-280.

[41] H. Saito "Hybrid Connection Admission Control in ATM Networks", Proceedings of the ICC 1992, pp. 699-703.

[42] H. Saito, K. Shiamoto "Dynamic Call Admission Control in ATM networks", IEEE Journal on Selected Areas in Communications, Vol. 9, No 7, Sept. 1991, pp. 982-989.

[43] H. Saito "New Dimensioning Concepts for ATM Networks", 7th ITC Seminar, Morristown, 1990.

[44] P. Sen, B. Maglaris, N. Rikli, D. Anastasiou "Models for Packet Switching of Variable-bit-rate Video sources", IEEE JSAC, Vol. 7, June 1989, pp. 834-843.

[45] M. Sidi, W. Liu, I. Cidon and I. Gopal "Congestion Control through Input Rate Regulation," proceedings of Globecom'89, Dallas, Texas, 1989, pp. 1764-1768.

[46] K. Sriram, W. Whitt "Characterizing Superposition Arrival Processes in Packet Multiplexors for Voice and Data", IEEE JSAC, Sept. 1986, pp. 833-846.

[47] T. Stern, A. Elwalid, "Analysis of Separable Markov-Modulated Rate Models for Information-Handling Systems", Advances in Applied probability 23, 1991.

[48] L. Trajkovic, S. Golestani "Congestion Control for Multi-Media Services", IEEE Networks Magazine, Vol. 6, No. 5, Sept. 1992, pp. 20-26.

[49] J. Turner "New Directions in Communications (or Which Way to the Information Age?)", IEEE Communications Magazine, Vol. 24, No. 10, Oct. 1986, pp. 8-15.

[50] J. Turner "Bandwidth Management in ATM Networks Using Fast Buffer Reservation", IEEE Networks Magazine, Vol. 6, No. 5, Sept. 1992, pp. 50-58.

[51] L. Zhang "Virtual Clock: A New Traffic Control Algorithm for Packet Switching Networks," Proc. Sigcomm'90, Philadelphia, Sept. 1990, pp. 19-29.

[52] D. Verma, D. Ferrari "Variation of Traffic Parameters in ATM Networks" Proceedings of ICC 1992, pp. 689-693.