In Proc. of the ITC Sponsored St. Petersburg International Teletraffic Seminar, June 1995

Traffic Controls to Support New Telecommunication Services Based on ATM: A Review of Activities of the ITU and ATM Forum

Arthur W. Berger

AT&T Bell Laboratories, Room 1K-211, 101 Crawfords Corner Rd, Holmdel NJ, 07733, USA

Abstract

The International Telecommunications Union and the ATM Forum are defining new ATM-layer bearer capabilities and associated traffic-control mechanisms. These bearer capabilities will be used by service providers for the design of telecommunication services based on ATM. The present article reviews and compares these bearer capabilities.

1. INTRODUCTION

Asynchronous Transfer Mode (ATM) offers the potential of supporting a wide range of existing and new telecommunication services. To realize this potential, one of the challenges is to design traffic controls that enable both economically efficient use of network resources and desired quality of service to higher layer applications. In order that such traffic controls can function across different networks and function across equipment from different manufacturers, at least some components of such traffic controls need to be standardized. Towards this objective, the International Telecommunications Union, in Question 8 of Study Group 13 has been active in preparing a new release of the Recommendation I.371 on "Traffic Controls and Congestion Controls in B-ISDN." Likewise, at the ATM Forum, the Traffic Management Working Group is preparing Version 4.0 of the "Traffic Management Specification." The present paper includes results from the 6 - 10 February meeting of the ATM Forum, [1], and the 27 February - 7 March 1995 meeting of Question 8 of the ITU, [2].

To put into perspective the task of designing suitable traffic controls, consider the crude solution where each connection specifies a peak cell rate and the network operator allocates bandwidth equal to that peak cell rate and checks that the traffic on the given connection does not exceed the stated peak cell rate. Such a solution essentially limits the role of ATM to bit-pipes, or circuits, of roughly any rate. Although this is still a significant role, it is far smaller than the full potential for ATM. For sources that are naturally bursty the above solution is clearly awkward: the network resources are poorly utilized and the resulting costs would be high and not competitive with other possible alternatives. Both the ITU and the ATM Forum have chosen to define a set of ATM layer bearer capabilities, which have different traffic control schemes associated with them. (In the following the term "bearer capability" is used as a shorthand for "ATM-layer bear capability.") Although there is some overlap in the features of the bearer capabilities and more than one of the bearer capabilities may be suitable for a given application, the set of all bearer capabilities is intended to be able to support effectively a wide range of applications, such that ATM could become the technology of choice for many present and new telecommunications services.

In the ITU, four ATM-layer bearer capabilities are defined: (1) Deterministic Bit Rate bearer capability, (2) Statistical Bit Rate bearer capability, (3) Available Bit Rate bearer capability, and (4) ATM Block Transfer bearer capability. In addition, a fifth bearer capability is under study, called the Unspecified Bit Rate bearer capability. The ATM Forum uses the term "service class" for "ATM-layer bearer capability" and defines service classes that correspond to first three bearer capabilities of the ITU as well as the Unspecified Bit Rate service class. (Also, the ATM Forum uses the term "Constant Bit Rate" instead of "Deterministic Bit Rate," and the term "Variable Bit Rate" instead of "Statistical Bit Rate.") These bearer capabilities contain standardized components of traffic control schemes, which the network operator can use in an overall customized control scheme.

The paper reviews and compares the five ATM-Layer bearer capabilities. The following section contains their description, and Section 3 contains a comparison of the bearer capabilities from various perspectives. A brief summary is then given in Section 4.

2. DESCRIPTION OF ATM-LAYER BEARER CAPABILITIES

2.1 Deterministic Bit Rate Bearer Capability

The Deterministic Bit Rate (DBR) bearer capability is used by connections that request a static volume of bandwidth, which is available at any time during the lifetime of the connection. The volume of bandwidth is specified by the Peak Cell Rate (PCR). (At public interfaces, the PCR is specified with a tolerance, known as the Cell Delay Variation (CDV) tolerance. The specification is made precise via an algorithm called the Generic Cell Rate Algorithm, where the PCR and the CDV tolerance are parameters in the algorithm. See [1] or [2] for details.) In the DBR bearer capability, given that the user's submitted traffic is conforming to the specified PCR, the network operator commits to a Quality of Service that includes a specified cell-loss-ratio objective and an end-to-end cell-delay-variation objective sufficient to support an application relying on circuit emulation. In the DBR bearer capability the source can emit cells at the PCR at any time and for any duration and the QoS commitments still pertain.

An obvious application for the DBR bearer capability is to support circuit emulation at a higher layer. However, this is not the only application for DBR. With circuit emulation, from the viewpoint of the ATM layer, there is an ongoing stream of cells, nominally spaced at the reciprocal of the PCR. In contrast, another type of ATM connection that the DBR bearer capability supports is one where the source may emit cells with nominal spacing equal to or greater than the reciprocal of the PCR, and may even emit no cells for periods. An important example is semi-permanent connections that may be established for periods of months but whose usage, at the ATM layer, varies during the course of the day. An important special case is a semi-permanent user-to-user Virtual Path Connection that is established between two locations of a corporation. During periods when the source is not making full use of the connection, the network operator can use the momentarily unused capacity for traffic from other connections.

The traffic controls used in the DBR bearer capability include the negotiation of the PCR and the subsequent policing of the submitted cell flow for conformance to the said PCR. To obtain the QoS commitment of a tight end-to-end cell delay variation, the network operator may need to isolate the DBR connections from other types of connections. One way network nodes could attain this isolation, as well as attain the feature of non-DBR connections using momentarily spare capacity of DBR connections, is to serve the DBR connections with higher priority than other types of connections.

2.2 Statistical Bit Rate Bearer Capability

In the Statistical Bit Rate (SBR) Bearer Capability, the end-system uses standardized traffic parameters to describe, in greater detail than just the Peak Cell Rate, the cell flow that will be emitted on the connection. A source describes the traffic to be emitted on the connection via the standardized traffic parameters: Peak Cell Rate, Sustainable Cell Rate (SCR), and Intrinsic Burst Tolerance (IBT), as well as the associated tolerances at public interfaces. The source may also describe the traffic via the declaration of the "Service Type" that pertains for the connection, [2]. The pair of parameters SCR and IBT are specified via the Generic Cell Rate Algorithm, where the reciprocal of the SCR is the increment and the IBT is the threshold, see [1] or [2] for details. The SCR is always less than the PCR. The traffic parameters are (typically) static for the duration of the connection, though re-negotiation of parameter values, via signalling or management procedures, will be possible in the future. In SBR the SCR traffic parameter is at the cell level in contrast to the ATM Block Transfer bearer capability where the SCR is at the block level.

At the ATM Forum, the SBR bearer capability includes a significant role for the Cell Loss Priority (CLP) bit. The CLP bit distinguishes between high priority (CLP=0) cells and low priority (CLP=1) cells of a connection. The CLP bit can be set by the source, or as an option the network may provide the "tagging" feature, whereby, at the request of the user, for cells submitted with CLP=0 that are not conforming to a traffic parameter for the CLP=0 cells, the network changes the CLP bit to 1 and tests the cell for conformance as a CLP=1 cell.

The following role of the CLP bit in the SBR (a.k.a. Variable Bit Rate) bearer capability is based on the ATM Forum Specification issued in 1993, [3], and is also included in the expanded Traffic Management specification, [1]. The following has also been proposed at the ITU, but is currently not agreed to. In the SBR bearer capability, the following configurations of the PCR and SCR/IBT traffic parameters with the CLP bit and the tagging option are specified:

1. The PCR traffic parameter and the SCR/IBT traffic parameters for the aggregate of all cells, regardless of the value of the CLP.
2. The PCR traffic parameter for the aggregate of all cells, and the SCR/IBT traffic parameters for the CLP=0 cells.
3. The same as (2.) above except with the tagging option invoked.

When the connection has both CLP=0 and CLP=1 cells, the cells entering the network with CLP=0 can be viewed as the committed portion of the traffic (the QoS commitment includes a specified cell loss ratio for the CLP=0 cells) and the cells entering the network with CLP=1 (or cells tagged by the network as CLP=1) can be viewed as the "at risk portion" (the cell loss ratio is Unspecified for the CLP=1 cells).

The SBR bearer capability can be partitioned into real-time and non-real-time. In the ATM Forum document, [1], the real-time VBR Service Class (SBR Bearer Capability) is for applications "requiring tightly constrained delay and delay variation, as would be appropriate for voice and video applications." Non-real-time VBR makes no commitment on delay variation, but does include a commitment on cell transfer delay: Whether this commitment is in terms of a mean or a maximum cell transfer delay is currently for further study. At the ITU, the use of the SBR bearer capability for real-time applications is for further study.

As with DBR, key traffic controls used in the SBR bearer capability include the negotiation of the traffic parameters PCR and SCR/IBT or the declaration of the Service Type and the subsequent policing of the submitted cell flow. In addition, the connection admission control (CAC) policy used in SBR enables the network operator to attain a statistical multiplexing gain, as compared with a CAC policy that allocates the PCR to the connections, while still meeting QoS commitments. Many such CAC policies are possible and the choice is at the discretion of the network operator. A network operator could choose a conservative policy that is based on the "worst case" traffic allowed by the PCR and SCR/IBT traffic parameters of the new and the established connections. Less conservative policies could use measurements of traffic of currently established connections and/or historical measurements of previous SBR connections. The historical measurements could be used to determine and/or validate stochastic models of the source traffic, and these models in turn could be used by the CAC policy.

2.3 Available Bit Rate Bearer Capability

The Available Bit Rate (ABR) bearer capability has been the focus of intense interest at the Traffic Management working group of the ATM Forum and to a lesser extent at the Resource Management Rapporteur's group of Study Group 13 at the ITU. ABR is designed for applications that can adapt their information transfer rate based on feedback information from the network; the feedback information is conveyed in Resource Management (RM) cells and provides an indication of the currently available bandwidth for the connection. Thus, real-time traffic controls are an inherent feature of this bearer capability. Much of the discussion over the past year and a half at the Traffic Management working group of the ATM Forum has been on the specifics of the traffic controls that support ABR, [1]. The text on ABR at the ITU can be viewed as a generalization of the text of the ATM Forum.

In ABR, a source will specify a PCR and a minimum cell rate, where the latter may be zero. The bandwidth available from the network may become as small as the minimum cell rate. The network commits to a QoS specified on the cell loss ratio, given that the cell flow is conforming at public interfaces, which would be the case if the source and destination follow a specified reference behavior, and certain delay bounds between source and interface are met. The network also makes the relative assurance that for connections that share the same path, no connection shall be arbitrarily discriminated against nor favored.

As a rough summary of the ABR control procedures, a source determines the current bandwidth provided for the connection by periodically emitting RM cells, some of whose fields may get modified at network nodes to indicate to the source a revised allowed rate. When the destination receives the RM cell, the destination sends the RM cell back to the source (on the companion connection in the reverse direction). When the source receives the RM cell, the source updates the allowed sending rate, based on the information in the RM cell. The ATM Forum has agreed to many additional aspects of the above mechanism, see [1] for details. Although the mechanism in [1] may seem complicated, one should bear in mind that the traffic control for ABR is being designed to function in both the public and private environment, over a great range of possible distances and speeds, and through network nodes handling various numbers of connections and with differing capabilities.

2.4 Unspecified Bit Rate Bearer Capability

The Unspecified Bit Rate (UBR) bearer capability is intended for delay tolerant applications. In contrast to ABR, UBR does not use a feedback traffic-control mechanism; however, such mechanisms may be operating at a layer above the ATM layer. Although the network operator may engineer resources to support UBR connections, the specification of UBR does not include QoS commitments on cell loss or cell delay. UBR can be viewed as a simple cell relay service analogous to the common term "best effort service." The PCR traffic parameter is negotiated for a UBR connection as it may identify a physical bandwidth limitation of the application or of a link along the path of the connections, [1]. Network nodes that support UBR connections would need to isolate in some fashion the non-UBR connections (if present) from the UBR connections. In addition, a desirable feature would be some isolation of each UBR connection from the other UBR connections. This could be accomplished via queueing and buffer management schemes.

2.5 ATM Block Transfer Bearer Capability

The ATM Block Transfer (ABT) bearer capability introduces the concept of a block of cells and, with use of RM cells, transports complete blocks with low cell loss and cell delay variation for cells of a block, comparable to that from the DBR bearer capability. "An ATM block is a group of cells delineated by two Resource Management cells, one before the first cell of the block and another after the last cell of the block," [3]. In a typical case, a higher layer protocol data unit would be packaged as an ATM block, though the use of ATM blocks is not restricted to this case. In ABT at connection establishment no bandwidth is allocated to the connection. Rather bandwidth is subsequently allocated on a block by block basis. Two ABT bearer capabilities are defined. In ABT with delayed transmission (ABT/DT) the source sends an RM cell to request a rate at which to transmit a block and then the source waits for a response RM cell from the network before sending the rest of the block. In ABT with immediate transmission (ABT/IT), "a user wishing to transmit an ATM block sends an RM cell immediately followed by the remaining part of the ATM block," [3]. If a network node along the connection can not support the requested rate, the request is denied, and in the case of ABT/IT the cells of the block may be discarded. At connection establishment the user may negotiate a sustainable cell rate traffic descriptor to obtain a guaranteed bandwidth for the connection.

The network operator engineers the resources and implements a connection admission control policy to keep within a specified level the probability a request is denied. "In the ABT framework, resource allocation within the network is block oriented, that is resources needed for the transfer of an ATM block are dynamically allocated on a block basis," [3].

2.6 Summary of Key Features

The following table is based on the table in [1] and expanded to include the ABT in [3].

Table 1
Summary of ATM-Layer Bearer Capabilities

	ATM-Layer Bearer Capability:					
Attribute	DBR	SBR real-time	SBR non-real-time	ABR	UBR	ABT
Cell loss ratio	specified, note 1			specified note 2	unspecified	specified note 3
Cell Transfer Delay and Delay Variation	specified		specified note 4	unspecified note 5	unspecified	specified
Peak Cell Rate	specified			specified note 6	specified note 7	specified
SCR/IBT	not applicable	specified		not applicable		specified note 8
Real-time Control via RM cells	no			yes	no	yes

Note 1 For DBR and SBR, the cell loss ratio may be unspecified for CLP=1 cells.
Note 2 Cell loss ratio is minimized for sources that adjust the cell flow in response to control information.
Note 3 For conforming blocks, the cell loss ratio is comparable to that for DBR. Block loss ratio is specified.
Note 4 For non-real-time SBR, the cell transfer delay is specified and the end-to-end cell delay variation is unspecified.
Note 5 The objective of the service is that the network does not excessively delay the admitted cells.
Note 6 Represents the maximum rate at which the source is allowed to send.
Note 7 Peak Cell Rate is not subject to CAC and UPC procedures.
Note 8 SCR specified for ATM Blocks.

3. COMPARISON OF ATM-LAYER BEARER CAPABILITIES

3.1 Cost

When designing telecommunication services based on the above bearer capabilities an obvious factor is the price to the user. If, hypothetically, the price were not an issue, then a user would choose a DBR connection at the desired peak cell rate, and the other bearer capabilities would be irrelevant. Since price is indeed relevant, one can view the non-DBR bearer capabilities as targeting a trade-off between some degradation in performance in exchange for cost savings, relative to DBR. The various non-DBR bearer capabilities can be viewed, conceptually, as different points on a plot of performance versus cost, where the particular features of a given bearer capability are intended to be well suited for some class of applications. An obvious point is that when a user has a choice between services based on these various bearer capabilities, the strategic choices by the service provider on the pricing scheme for a given service will be a significant factor in the relative success of the service. Thus, the statement that users will choose bearer capability "A" for Application "X" is actually conditioned on the pricing schemes of various alternatives.

3.2 Target Applications

The Traffic Management working group of the ATM Forum lists the target applications for each bearer capability, [1]. DBR is "intended for real-time applications, i.e. those requiring tightly constrained delay and delay variation, as would be appropriate for voice and video applications." Real-time SBR (VBR) can be used "for any CBR (SBR) application for which the user can benefit from statistical multiplexing by sending at a variable rate and can tolerate or recover from the resulting small but non-zero random lost rate." Non-real-time SBR (VBR) can be used for transaction processing such as airline reservations and for Frame Relay interworking. UBR is targeted for computer communications, and ABR is targeted for LAN emulation and "any UBR application for which the user wants a more reliable service." ABT is targeted for the same applications as non-real-time SBR and as ABR, where, as in SBR, the user negotiates a SCR and as in ABR the allowed bandwidth varies during the lifetime of the connection, where the user requests the bandwidth, as opposed to being told the available bandwidth as in ABR.

3.3 Information at Connection Establishment

In the DBR and SBR bearer capabilities, the user needs to specify values for the traffic descriptors at connection establishment and the subsequent realized traffic flow is suppose to conform to these traffic descriptors. In the ABR and ABT bearer capabilities, although the user also specifies a priori values for traffic descriptors, the realized traffic flow in the case of ABR adapts to real-time feedback from the network and in the case of ABT conforms to the peak cell rate that has been most recently requested. Providing the a priori information required for DBR and SBR may or may not be convenient for a given application. For example, consider the application of retrieval of digital image information, such as medical imaging from a remote location. Given some knowledge of the size of the images and desired response times, the SBR bearer capability could be very appropriate. Suppose that the number of bytes needed to encode the image is less than some number of Megabytes. Suppose that one of the desired QoS commitments is that the delay from the time epoch a new image is requested to the epoch when the new image is fully displayed is less than d seconds. Lastly, suppose the time interval between requests for a new image is more than e seconds, where e is greater than d. With this information, the traffic parameters PCR and SCR/IBT can be chosen for the connection. On the other hand, if such information is not known and the delay requirements are somewhat soft, then the ABR bearer capability could be very appropriate. (Again, the pricing scheme of the telecommunication service will be a relevant factor.)

Representatives from the computer industry have argued that application software would not know what SCR/IBT to pick or know a current PCR to request for an ATM block. They have argued for the ABR bearer capability as data applications could adapt to the varying bandwidth of ABR via the provided feedback. A somewhat contrasting perspective is provided by the current example of Frame Relay service where the user needs to pick a priori a committed information rate and a committed burst size, which are analogous to the SCR/IBT, where the Frame Relay traffic may be an aggregation of traffic from higher layer applications. Even though customers do have some difficulty in choosing values for these parameters, the Frame Relay service has been successful, in part, as an attractive alternative to private lines due to cost savings realized by the customer.

3.4 Installed base of Customer Premise Equipment

Although private ATM networks will be deployed early-on by some businesses, public service providers should account for the large installed base of non-ATM customer premise equipment. An important early, as well as ongoing, application of the ATM-layer bearer capabilities is actually not as a new telecommunications service but rather as a technology replacement within the public network. For example, a public carrier could provide Frame Relay Service over an ATM network that uses the SBR bearer capability. Of course, over time, both the customer premise equipment and the service offerings by public carriers evolve, in part, in reaction to the other. Any ATM telecommunications service, by definition, has an ATM interface with the pubic network and thus presumes that the customer has installed some ATM equipment. In the case of a business customer this ATM equipment may initially be a gateway that interfaces between the ATM traffic to/from the public network and the non-ATM traffic in the interior of the customer premise. An early telecommunication service could be based on DBR where the customer views the service as an alternative to private lines, with the new feature of greater choice of bandwidths. Likewise UBR may be suitable in the case of simpler customer premise ATM equipment or existing application software where neither makes use of the feedback information provided by ABR. In the USA, there is strong interest in the ABR and SBR bearer capabilities, and one can expect that some service providers will offer follow-on services based on these bearer capabilities.

4. SUMMARY

Collectively, the ITU and the ATM Forum are defining five ATM-Layer Bearer Capabilities: (1) Deterministic Bit Rate bearer capability, (2) Statistical Bit Rate bearer capability, (3) Available Bit Rate bearer capability, and (4) ATM Block

Transfer bearer capability, and (5) Unspecified Bit Rate bearer capability. Although the bearer capabilities have distinct attributes, they overlap considerably in the applications that they could potentially support. Service providers can make strategic choices in the use of the bearer capabilities and of price positioning for the design of new telecommunication services

1. "Draft Version 2.0 of ATM Forum Traffic Management Specification Version 4.0," ATM Forum Technical Committee, Document number atmf 95-0013R1, February 23, 1995.
2. "Traffic Control and Congestion Control in B-ISDN," ITU-T Draft Recommendation I.371, Paris, France, 27 February - 7 March 1995.
3. Living List for ITU-T Draft Recommendation I.371, Paris, France, 27 February - 7 March 1995.
4. ATM User-Network Interface Specification, Version 2.4, ATM Forum, August, 1993.