PREVENTIVE METHODS FOR ATM MODE CONTROL

Ivan Baroňák, Róbert Trška

Department of Telecommunication, FEI STU, Bratislava

Summary Broadband transfer mode ATM represents one of alternative solutions for growing requirements on transfer capabilities. Its advantage is an effort for provision of guaranteed quality of transport services with preservation of high transfer rate. This property is covered by several mechanisms, which role is to control not only the traffic of existing connections, but also the admission of new ones and prevent the violation of requirements on transport quality of existing and new connections.

Keywords: ATM, Connection/Call Admission Control (CAC), Quality of Services (QoS)

1. INTRODUCTION

Expansion of information technologies in last years forced providers of transport networks to look for new transport technologies able to cover the growing needs of customers.

Asynchronous Transfer Mode (ATM) started up as one of possible solutions with suitable requirements on fast and reliable transfer of huge amounts of data on various distances. It combines the packet-oriented data transfer with transfer on predefined transmission links. ATM is the universal technology, which provides transport services regardless of type of terminal, service or requirements on transfer rate. Its clear advantage, especially compared to connectionless technologies, is the support of quality of services (QoS), which enables the user to specify the transport requirements (e.g. the transfer rate, transfer delay, etc.) before the connection is established and when the network accepts these requirements, they can be guaranteed during the whole connection duration.

2. ATM TRAFFIC CLASSIFICATION

To enable the transport of various data types in ATM network in accordance with user's requirements on QoS, several mechanisms exist in ATM and their role is to guarantee the negotiated QoS.

One of these fundamental mechanisms is the classification of traffic into several categories following the requirements on transfer rate, timing, etc. Within the ITU-T (Internation Telecommunication Union – Telecommunication Standardization Sector) there was created the distribution of services into four service classes:

Class A – supports the circuit emulation and is assigned for transport of audio and video signal with constant transfer rate.
Class B – is suitable for transport of audio and video signal with varying transfer rate.
Class C – serves for connection-oriented data transport.
Class D – is suitable for connectionless data transport.

To these service classes following service categories are equivalent, which were approved by ATM Forum:

Constant Bit Rate (CBR) Service – requires constant bandwidth during the whole connection duration and is assigned for real-time applications with time transparency (voice, audio, video, etc.).
Real-time Variable Bit Rate (rt-VBR) Service – is suitable for applications requiring variable bandwidth and time transparency and for audio and video applications with variable transfer rate.
Non-real-time Variable Bit Rate (nrt-VBR) Service – is assigned for applications, which does not require real-time transfer or constant transfer rate.
Unspecified (Unassigned) Bit Rate (UBR) Service – has not specified the transfer rate, doesn't guarantee the QoS and is used for applications, which does not require time transparency (e.g. the file transport). It is the "best effort" service.
Available Bit Rate (ABR) Service – uses the unexploited bandwidth in multiplex and is useful for data transport.

ITU-T also defined the set of ATM Transfer Capabilities (ATC), which specify the model of services built from traffic characteristic and requirement on QoS.

3. QoS AND TRAFFIC PARAMETERS

In the initial phase of connection establishment in ATM network it is necessary to negotiate the characteristic properties of required service between the user and the network, for example the type of service, traffic parameters, QoS requirements for both connection directions, etc. These parameters together form the traffic descriptor of this connection. The character of traffic on the connection is defined by traffic parameters; the set of QoS parameters specifies the user requirements on QoS.

Peak Cell Rate (PCR) – specifies the maximal rate of traffic source.
Minimum Cell Rate (MCR) – minimal transport bandwidth required by network.
Sustainable Cell Rate (SCR) – represents the average transfer rate in adequate time interval.

Burst Tolerance (BT) – defines the upper boundary for burst length, which the network guarantees to transmit.
Maximum Burst Size (MBS) – specifies the maximal number of cell, which can be transmitted on PCR, while the SCR must be preserved.
Peak-to-peak Cell Delay Variation (CDV) – variable component of transfer delay (jitter).
Cell Delay Variation Tolerance (CDVT) – the boundary for CDV.
Cell Transfer Delay (CTD) – time interval between two successive successful cell transfers.
Cell Loss Ratio (CLR) – number of lost cells to the number of all cells.

Not all these parameters must be used for every service type.

4. TRAFFIC CONTROL IN ATM

As the ATM protocol is based on principle of statistical multiplexing, it comes from the assumption, that not every connection will utilize full capacity of transmission lines and therefore it is possible to more effectively exploit the assigned capacity of these lines. In the process of connection negotiation the function of Connection/Call Admission Control (CAC) is executed and when new connection requests to enter the multiplex, the CAC function determines, whether there is sufficient capacity of network resources to preserve the negotiated QoS of existing connections and based on this it decides, whether the new connection will be accepted or not. CAC is the preventive mechanism, it is the primary protection of network from congestion and its role is to limit the number of traffic flows in the network so every flow keeps at disposition the requested QoS.

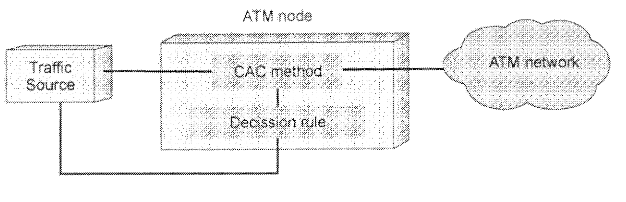

Fig. 1. An example of CAC realization in ATM network

Despite of the CAC function it can come to situation, when more traffic sources begin to transmit on its peak cell rate and the result is the congestion in network, the risk of increased cell loss and consequently the degradation of quality of connection. As ATM doesn't include any mechanism for retransmission of cells lost due to the congestion, the attention must be given especially to the minimizing of the possibility of cell loss and to the increase of network throughput. Functions belonging to reactive mechanisms serve to this purpose. They monitor traffic flows and in the case of imminent or existing congestion they perform relevant actions. There exist more solutions of congestion problem in network: the use of buffers, packet discard, feedback control, etc.

Next we will focus only on preventive mechanisms for traffic control.

5. CAC METHODS

CAC methods are the most important component of traffic control in ATM networks. Their task is to keep the balance between two opposing requirements: between the maximizing of network resources utilization and the observance of negotiated parameters of connection. They are the first proceeding in the process of allocation of network resources to VP/VC connection. The CAC method is a decision-making algorithm, which on the arrival of request for new connection establishment should decide, whether to permit or refuse the new connection regarding the given level of network resources utilization. New connection can be created only when there are sufficient resources able to meet the QoS requirements of new connection, so the requirements of existing connections remain unbroken.

In the case, when the new connection is accepted, the traffic contract between the user and the network is made. In this contract the network commits to keep the negotiated QoS parameters on the assumption that the user keeps his duties resulting from the contract.

For the CBR and VBR services is CAC the obligatory function of preventive traffic management. The ABR service use the feedback control as the QoS mechanism and the UBR service has no network provided QoS neither reserved network resources.

In contrast to deterministic multiplexing, which uses the allocation of transmission lines based on the peak cell rate, the statistical multiplexing uses the effective bandwidth, which is in the case of VBR connections much lower then the peak cell rate and thus it allows more efficient utilization of network resources. However, the effective implementation of statistical multiplexing is more problematic, because it needs the detailed knowledge of connection behavior and it must be able to return meaningful results in real-time process. Good CAC method should fulfill the following criteria:

Simple – the implementation of the method should be economical and fast.
Fast – able to generate real-time results.
Flexible – adaptable with respect to new services.
Robustness – effectively working also with not fulfilled assumptions.
Efficient – maximize the utilization of statistical multiplexing.
Effective – able to guarantee the negotiated QoS.
Controllable – allows traffic control without degradation of network performance.

The main part of CAC algorithms is the estimation of bandwidth requested by connection set so the QoS requirements of every connection will be fulfilled. In other words, it is the solution of the problem, when for N multiplexed connections with bit rates $r_i(t)$ the probability that the instantaneous aggregated bit rate exceeds the total required bandwidth C is lower then the given value ε. This probability can be expressed as

$$P\left[\left(\sum_{i=1}^{N} r_i(t)\right) \geq C\right] < \varepsilon .$$

CAC method in the case of CBR connections is in principle trivial. The main study is therefore oriented to CAC methods for VBR traffic, where several different approaches exist.

5.1. Gaussian approximation method

This method for estimation of required bandwidth is quite fast and simple. It goes from Gaussian distribution, which is used for distribution approximation of aggregated traffic providing the large number of connections and none connection is dominating. Each connection is characterized by two parameters – the mean arrival rate λ_i and the standard deviation σ_i. The total mean arrival rate λ and the variance of arrival rate σ^2 are defined as the sum of values of individual connections. The estimation of buffer overflow probability is defined as

$$P(overflow) = P\left[\left(\sum_{i=1}^{N} r_i(t)\right) \geq C\right] = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\lambda-C)^2}{2\sigma^2}}$$

and the upper bound for cell loss probability is

$$P(loss) = \frac{E\left[\left(\sum_{i=1}^{N} r_i(t)\right) - C\right]^{+}}{\lambda} \leq \frac{\sigma}{\lambda\sqrt{2\pi}} e^{-\frac{(\lambda-C)^2}{2\sigma^2}} ,$$

where $r_i(t)$ is the instantaneous bit transfer rate of connection i.

This method achieves good results only for large number of connections. Moreover, all connections are considered as equivalent in cell loss requirements and we consider the bufferless system model. The Gaussian method cannot fully exploit the statistical multiplexing gain.

5.2. Convolution method

This method is also based on queuing model without buffers and is considered as very exact. It uses simple model of burst traffic with peak rate R and mean rate r. The probability that the connection is in the burst state is r/R and in the idle state $1 - r/R$. The connection holding time distribution is arbitrary. For modeling of individual connections the fluid flow traffic model is used. Total bandwidth requirements are calculated using convolution of cell rate distribution of each connection. This model allows exact cell loss estimation, but the computation requirements grow quickly with increasing number of connection and therefore it is necessary to use some approximations.

5.3. Effective bandwidth method

Using this method we can determine the effective bandwidth, which is between the peak rate and mean rate of connection. In contrast to foregoing methods, the effective bandwidth method supports the cell loss bound for each connection and supports also the buffers. In the case, when we assume large buffers, we can approximate the cell loss rate with exponential function depending on buffer size.

Effective bandwidth c of single source corresponding to upper bound of cell loss rate ε can be derived using fluid flow model and also using the on-off model of traffic. Exact solutions are very hard for computation and therefore some approximations are used in praxis. One of them assumes fluid flow traffic model ignoring the effect of multiplexing gain, what can cause the overestimation of required bandwidth. Another approximation is Gaussian approximation (large N, bufferless model) and it gives lower bandwidth requirement as assumed. The required bandwidth is then given as minimum value from $\{C_g, C_e\}$, where

$$C_g = \sum_{i=1}^{N} \lambda_i + \left(\sqrt{-2\ln(\varepsilon) - \ln(2\pi)} \sum_{i=1}^{N} \sigma_i\right),$$

$$C_e = \sum_{i=1}^{N} c_i .$$

5.4. Diffusion based method

The diffusion based method is based on two formulas, which describes required bandwidth for diffusion models with finite and infinite capacity queuing system, respectively using the instantaneous return process:

$$C_{df1} = \lambda - \delta + \sqrt{\sigma^2 - 2\sigma^2\omega_1} ,$$

$$C_{df1} = \lambda - \delta + \sqrt{\sigma^2 - 2\sigma^2\omega_1} ,$$

where

$$\omega_1 = \ln\left(\varepsilon\sqrt{2\pi}\right),$$

$$\omega_2 = \ln\left(\varepsilon\lambda\sqrt{2\pi}\right) - \ln(\sigma),$$

$$\delta = \frac{2B}{\kappa}\delta^2 ,$$

ε is the maximum acceptable cell loss rate, κ is the instantaneous variance of the cell arrival process, and B is the size of buffers.

The use of two statistical bandwidths consider the interaction between individual streams in ATM multiplexer with respect on user's cell loss requirements, characteristics of aggregated traffic and available buffer space in multiplexer. They also define the respectable range for various connection classes based on their traffic descriptors for various scales of buffers.

The diffusion based method is more conservative with respect to cell loss, but more economical in bandwidth allocation and it offers larger admission regions for homogenous and heterogeneous traffic.

5.5. Fast buffer reservation method

The fast buffer reservation method uses the open loop scheme. The purpose of this method is the preservation of integrity of bursts as a whole. This mechanism is suitable for bandwidth reservation on burst level, in contrast with previous methods, which works especially on the cell level.

Each virtual circuit passing through the switch is considered as a two-state machine, which can be active or idle. In the active state the source transmits on peak rate and predefined number of buffer slots equivalent to its peak rate is allocated to the active virtual circuit. It has guaranteed access to this buffer space until it returns to idle state. The transition between these two states is done on arrival of cells indicating the start and the end of burst. In the case, when at the time of start-of-burst cell arrival there is not enough free buffer space available to hold the burst, the whole burst is immediately discarded.

5.6. Measurement method

The computational complexity of CAC methods caused, that also the approaches derived from practical experiences are used, although they not always meet the requirements and the application requesting for connection is not always able to characterize the traffic correctly before the transmission begins.

The measurement methods go from direct measuring of bandwidth requirements. It eliminates the need for specification of parameterized traffic model by new connection. The method relies on the on-line measurement of traffic going through the switch and it require only minimum of information from the new connection, e.g. the peak cell rate. Additional information may increase the efficiency of CAC process. The initial bandwidth estimation is calculated from available parameters and then it is adjusted according to measurement results. If the initial information is not available, more conservative estimation can be made from declared peak rate.

5.7. Artificial intelligence methods

The advantage of methods based on principles of artificial intelligence is their adaptability and the learning ability. The fuzzy logic systems and artificial neural networks belong to the most studied.

The field for fuzzy systems utilization is for example the situation, when the set of QoS parameters and traffic descriptors is small and incomplete and thus the right estimation of effective bandwidth and buffering requirements is difficult, or when the situations are varying in time and it is impossible or unpractical use exact mathematical models. On the other hand, the artificial neural networks are able to capture the complex nonlinear relationship and to estimate the model of arrival traffic or the QoS characteristic. Advantageous is their learning capability.

When good programmed and trained, these methods returns exact and fast results. The combination of profitable features of fuzzy systems and artificial neural networks seems to be the ideal solution. In such neural-fuzzy systems the inputs are led into system each time, when the request for new connection arrives and the system uses them for QoS prediction. The prediction allows the dynamic allocation of available buffer capacity in switches.

6. CONCLUSION

The ATM technology represents the possibility how to transport the large amounts of user information at high speed, with guaranteed quality of transmission and with respect to specific requirements of individual connection types. The preventive methods of traffic control like CAC play a significant role in quality of services provisioning. Many experts are studying hard these methods to improve the operating properties of individual methods. In addition to "classical" methods a notable potential is hidden in the utilization of methods based on principles of artificial intelligence.

REFERENCES

[1] Kajan, R.: Quality Of Services In ATM Networks, Telecommunications and Business, 10/1999, pp. 27 – 28

[2] Kilkki, K.: Traffic Characterisation and Connection Admission Control in ATM Networks, Disertation for the Degree of Doctor of Technology, Helsinki University of Technology, Espoo, 1994

[3] Kukura, P.: ISDN, B-ISDN, ATM – Digital Networks with Integrated Services, elfa, Košice, 2000

[4] Perros, H., Elsayed, K. M.: Call Admission Control Schemes: A Review, IEEE Communications Magazine, 11/ 1996, pp. 82 – 91

[5] Rahin, M. A., Kara, M.: Call Admission Control Algorithms in ATM Networks: A Performance Comparison and Research Directions, Interim Research Report Draft 0.8, School of Computer Studies, The University of Leeds, 1998