TRAFFIC MANAGEMENT IN ATM NETWORKS

Matej KAVACKÝ, Ivan BAROŇÁK
Slovak University of Technology Bratislava, Faculty of Electrical Engineering and Information Technology,
Department of Telecommunications, Ilkovičova 3, 812 19 Bratislava, Slovak Republic,
E-mail: matej.kavacky@stuba.sk, ivan.baronak@stuba.sk

SUMMARY

In this paper we present the traffic management functions in Asynchronous transfer mode (ATM) networks. In the first part of the paper commonly used traffic models for ATM traffic are discussed. Traffic management and its functions in the network are presented in the next part of the paper. Then connection admission control methods (CAC) as preventive traffic management mechanism are then presented. Overview of two CAC methods proposed in the literature from the groups of statistical based, measurement based and artificial intelligence based methods are presented in the final part of the paper.

Keywords: *ATM, traffic management, traffic models, connection admission control, artificial intelligence*

1. INTRODUCTION

One of the competitive advantages of Asynchronous Transfer Mode (ATM) technology is its capability of Quality of Servic (QoS) provision. This capability is very sophisticated in comparison with other types of networks. Hence, customer can choose and pay only the required service level. ATM technology allows concurrent effective transmission of voice, video, data and multimedia. Each of these services generates traffic with different characteristics that are modeled by various mathematical models for purposes of ATM traffic behaviour analysis and simulations. Miscellaneous models are employed, but the most used traffic models are:

Markov models are based on the discrete Markov chains using states to representation of bit rate modes. Markov property means that current state is dependent only on the previous state. These models are used to model On-Off sources. On-Off traffic source generates ATM cells at the peak rate in the On state and no cells are generated in the Off state. These models are widely used in voice and video modeling.

Fluid-flow models regard with the assumption that cell arrivals in discrete time are fluid flow of cells, so arrivals of particular data units are not considered. Advantage of fluid flow is its simplification of performance analysis. These models are especially suitable for the fast ATM networks, because intervals between arrivals of particular data units during active period are too short in comparison with rate changes so impact of particular arrivals are omissible from waiting queue performance analysis point of view. Fluid-flow models are suitable for voice and video sources.

Regression models explicitly define next random variable in sequence by previous variable within the scope of given time window and moving average of white noise. Autoregressive model of p^{th} grade, denoted as AR(p), is defined by following expression:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \varepsilon_t , \qquad (1)$$

where X_t denotes specified random variables, ε_t denotes white noise and ϕ_j are real numbers. Regression models are used to model sources with variable bit rates (e.g. video sources).

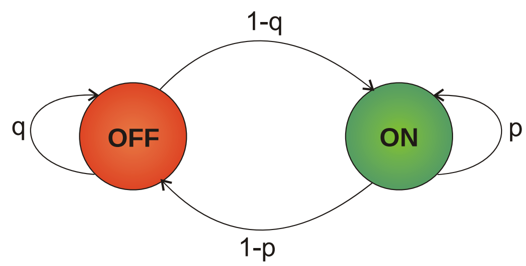

Fig. 1 Two-state Markov On-Off source model,
1-q: transition probability from Off to On state,
1-p: transition probability from On to Off state

Self-similar models are suitable for modeling of traffic incoming to ATM network from Local Area Networks (LAN). Ethernet LAN traffic measurements proved that this type of traffic has fractal behaviour, which could not be modeled by commonly used traffic models (e.g. Markov or Fluid-flow models). Ethernet traffic has structurally similar bursts on each level of time resolutions from seconds up to hours. This phenomenon is called self-similarity.

2. TRAFFIC MANAGEMENT

Implementation of mechanism of traffic control in ATM network – traffic management – is necessary requirement for providing declared quality of service. The main task of traffic management is to protect ATM networks and end terminals against congestion and also to achieve effective utilization

ISSN 1335-8243

of network resources. ATM Forum defines several basic functions in Traffic Management Specification that can be used for traffic and congestion control in ATM - Connection Admission Control, Usage/Network Parameter Control, Cell Loss Priority Control, Traffic Shaping, Frame Discard, Network Resource Management, Feedback Control. Traffic control can be preventive or reactive. ATM technology relies mainly on preventive control which is divided into two groups:

1) Admission Control which decides if a new connection can be accepted to the multiplex or it should be rejected. Among the functions defined by ATM Forum such functions are *Connection admission control* (CAC) and *Network resource management.*

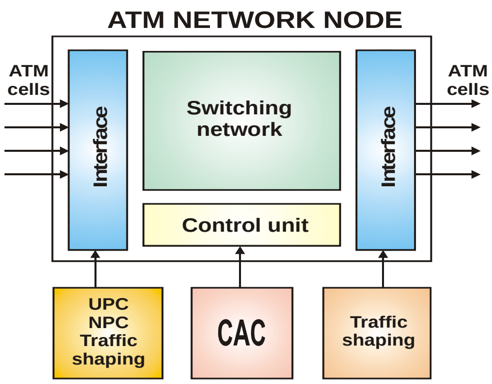

Fig. 2 Network node and traffic management functions

2) Police Function monitors traffic in multiplex in order to operative intervention in case of congestion state. Congestion can occur even if CAC function is applied when several sources with bursty traffic simultaneously increase transmission rate. Police functions are *Usage and Network Parameter Control (UPC/NPC), Cell Loss Priority Control (CLP), Traffic Shaping and Frame Discard*. Several mechanisms can be used for police functions, such as mechanism based on Leaky-bucket algorithm or various window functions. Mainly is used Leaky-bucket algorithm.

3. CONNECTION ADMISSION CONTROL

Connection admission control is preventive function that makes decision if a new connection is accepted to the ATM network or it should be rejected in order to maintain QoS of existing connections. Decision process is based on actual network load, values of characteristic parameters (e.g. peak and mean rates), availability of network resources, and required quality of service of existing and new connections. Requirements on quality of service are often expressed in terms of network performance parameters bounds – cell loss, delay variation and waiting queue delay. User expects that network will satisfy these parameter bounds. Another task of CAC methods is to optimize division of transmission bandwidth in order to maximize statistical gain of multiplexing. This gain is based on the assumption that not all multiplexed connections will require transmission at peak rate during whole duration of connection. Hence, available bandwidth can be assigned to new source that requests admission to the network, in the case if the QoS requirements of existing connection will be maintained.

Four main requirements should be considered during designing CAC methods [1]:

Effective bandwidth utilization and cell loss ratio CLR – better bandwidth utilization offers providing of cheaper service. One of the main task of CAC method is to achieve the most effective bandwidth utilization together with agreed CLR parameter maintenance. Effective bandwidth utilization is performance measure at network side; CLR is measure at user side.

Simple implementation – because CAC is real-time control procedure, reduction of processing complexity is necessary. Admission control must have simple implementation without complex hardware and software mechanisms.

Scalability and robustness – admission control methods should be effectively used for one connection as well as for several aggregated connections without waste increasing of processing complexity. CAC methods should be applicable for every traffic type.

Traffic model independence – CAC methods should not be dependent on specific traffic model, because it is limitation factor of their use. New service should not to raise need of new CAC method.

Although many CAC methods and their modifications have been developed, no single method is universal. Some of CAC methods are based on probability theory and statistics. Other CAC methods are based on on-line measurement of traffic. They use parametric traffic model (mostly On-Off model). Methods based on artificial intelligence (neural network, fuzzy logic) are also developed and extensively studied, because of their adaptive and learning capabilities. In the next part of the paper overview of few CAC methods from each group is presented.

3.1. Statistical methods

Statistical CAC methods are based on mathematical models and statistics. At the time of decision process these methods use statistical estimation of bandwidth required by set of connections so that quality of service of each connection will be preserved. Probability that instantaneous bit rate of N connections will exceed specified bandwidth C is lower than required cell loss probability is defined by following expression:

$$P\left[\left(\sum_{i=1}^{N} r_i(t)\right) \geq C\right] < \varepsilon \ , \qquad (2)$$

where N denotes number of multiplexed connections, C denotes required link capacity for these connections, $r_i(t)$ is instantaneous rate of i^{th} connection.

Convolution method is one of the most accurate statistical methods that can exactly determine distribution of aggregated bit rates on the ATM link. Method is based on convolution algorithm. If X is bandwidth required by new connection, Y is bandwidth required by existing connections and Q is total bandwidth required by all connections then we can calculate bandwidth Q by convolution:

$$Q = Y * X \ . \qquad (3)$$

This expression can be solved according to [10] by following equation:

$$P(Y + X = b) = \sum_{k=0}^{b} P(Y = b - k)P(X = k) \ , \qquad (4)$$

where b denotes bandwidth required by connections in given instant k.

Convolution method requires $N\text{-}1$ convolutions for each new connection. Equation (3) can be rewrited according to [5] in the form:

$$Q_n = Q_{n-1} * X_n ; n \in \{1,2,...,N-1\} \ , \qquad (5)$$

where N denotes number of existing connections and $Q_0=X_0$. It follows that $N\text{-}1$ convolutions are required for acquisition of whole distribution. Several disadvantages of convolution method are cost for its high accuracy:

- computation complexity in the case of higher number of connections and need of large memory,
- after connection termination deconvolution is not possible, new state vector [5] must be calculated from scratch,
- Quality of Service of particular connections is not considered.

Several modifications of convolution method have been proposed [6] that decrease its computation complexity, but on the other hand accuracy is also decreased.

Equivalent capacity method is also known as Effective bandwidth method. It is widely used and popular method. Decision if the set of connections will be accepted or rejected depends on the comparison of the sum of equivalent capacities of particular sources with total link capacity. Equivalent capacity is defined in [7] as a service rate that ensure required cell loss ε:

$$\varepsilon = \beta \exp\{-\frac{K(c - rR)}{b(1-r)(R-c)c}\} \ . \qquad (6)$$

Equivalent capacity can be solved from previous equation with simplification that $\beta=1$ (usually $\beta<1$). Then equivalent capacity c of one source is given as:

$$c = \frac{a - K + \sqrt{(a-K)^2 + 4Kar}}{2a} R \ , \qquad (7)$$

where R denotes peak rate of source, r denotes active period of source, b is mean time of active period of source, K denotes capacity of finite waiting queue and

$$a = \ln(\frac{1}{\varepsilon}) b(1-r)R \ . \qquad (8)$$

For N sources calculation of equivalent capacity is very complex so following approximation is used:

$$c = \min\{\rho + a'\sigma, \sum_{i=1}^{N} c_i\} \ , \qquad (9)$$

where c_i denotes equivalent capacity of i^{th} source and calculated according expression for one source, ρ is total average bit rate of particular sources.

Some studies showed that this method is inaccurate in some situations, for example in the case if the system without buffers generating input traffic has low probability that traffic will exceed link capacity. Equivalent capacity (Effective bandwidth) method supposes that this probability is close to 1. Some modifications of this method exist that improve its accuracy.

3.2. Measurement based methods

Measurement based CAC methods use online measurements in network to determine current traffic load instead of computation of traffic characteristics from user defined parameters and given traffic model. Traffic that passes through the ATM switch is measured and only little knowledge about new connection is required, often peak rate is enough.

Method based on UPC (Usage Parameter Control) parameters monitoring measures individual ATM cell flows. It monitors mean rate m and probability of cell tagging ξ in Leaky-bucket algorithm. Dynamic bandwidth allocation is used. If parameters (m, ξ) will exceed defined bounds, command for bandwidth increase or decrease is issued. Equivalent link capacity c with buffer of size K and cell loss probability ε constraint is given by the same expression as in the Equivalent capacity method, parameter r is given by the expression $r = m / R$, where R denotes source peak rate and m is the mean rate.

Decision if a new connection will be accepted or rejected is performed according the following equation:

$$r_j (= c) \le C - \sum_{n \ne j} r_n \ , \qquad (10)$$

where r_j denotes required bandwidth, r_n denotes bandwidth of particular existing connections, c denotes equivalent link capacity and C is total link capacity.

Method based on lowpass filter uses measurement of cell flow filtered by lowpass filter. Instantaneous rate is measured by lowpass filter and decision process is based on effective bandwidth derived from measured instantaneous rate. Effective bandwidth is set to the maximum instantaneous rate measured during monitoring period. Instantaneous rate can be estimated by application of recursive lowpass filter on monitored number of cells arriving during time slot according [4] by equation:

$$\lambda(t) = \alpha n(t) + (1-\alpha)\lambda(t-\delta) \ , \ 0 \le \alpha \le 1 \ , \qquad (11)$$

where *n(t)* denotes number of cells arrived during slot *t*, δ denotes time of cell transition along transmission link, α denotes smooth ratio and *λ(t)* is instantaneous rate of cells that arrived in time slot *t*.

Residual bandwidth is derived from difference between link capacity C and maximum instantaneous rate observed during sufficiently long time period. For a new connection with peak rate R is used following decision criteria [4]:

$$\frac{R}{C} < \rho - \max_{t' \in (t - T_m, t)} \lambda(t') \ , \qquad (12)$$

where T_m denotes monitoring interval, R denotes peak rate of new connection, C denotes link capacity and ρ is tolerance utilization ratio.

3.3. Methods based on artificial intelligence

The main idea of CAC methods based on artificial intelligence is to obtain relation between offered traffic and required quality of service by learning during on-line operation. Allocation of network resources for connections can be based on previous knowledge and can be adaptive towards changes in traffic characteristics. This principle is basis for methods based on artificial neural networks (ANN). Another group of CAC methods that share similar aims as learning principle of neural networks are methods based on fuzzy logic. Decision if new connection should be accepted or rejected is performed on the basis of relation between statistical behaviour of traffic and monitored CLR through model that is component part of system architecture and represents learnt knowledge. This model can be:

- ***set of neurons and interconnections between them*** in the case of artificial neural networks,
- ***set of „if – then“ rules***, that represents relations between input and output fuzzy variables, in the case of systems based on fuzzy logic.

Method based on estimation of equivalent capacity through fuzzy logic [2]. As the input traffic linguistic variables were selected connection peak rate R_p, connection mean rate R_m and time of peak rate duration T_p. As the output linguistic variable was selected estimated equivalent capacity denoted as C_e. On the basis of knowledge from conventional equivalent capacity method and computation of large number of equivalent capacities (10^5) for different combination of variables R_p, R_m and T_p, i.e. different traffic sources, we can say that set of three linguistic expressions is enough to describe peak rate R_p, to describe R_m is enough the set of two expressions and for describing T_p are three expressions enough. These sets of linguistic expressions for particular variables are according to [2] following:

- $T(R_p)$ = { Small (S), Medium (M), Large (L) }
- $T(R_m)$ = {Low (Lo), High (Hi) }
- $T(T_p)$ = { Short (Sh), Medium (Me), Long (Lg) }

Estimated equivalent capacity for connection should be in range between R_m and R_p. Proposed system divides this interval into six quantization levels. If C_i (i=1,..,6) denotes estimated equivalent capacity of i^th level, then set of expressions for output variable C_e is specified as $T(C_e) = \{C_1, C_2, C_3, C_4, C_5, C_6\}$.

Rule	R_p	R_m	T_p	C_e	*Rule*	R_p	R_m	T_p	C_e
1	S	Lo	Sh	C_1	10	M	Hi	Sh	C_1
2	S	Lo	Me	C_2	11	M	Hi	Me	C_2
3	S	Lo	Lg	C_5	12	M	Hi	Lg	C_5
4	S	Hi	Sh	C_1	13	L	Lo	Sh	C_4
5	S	Hi	Me	C_1	14	L	Lo	Me	C_6
6	S	Hi	Lg	C_4	15	L	Lo	Lg	C_6
7	M	Lo	Sh	C_1	16	L	Hi	Sh	C_3
8	M	Lo	Me	C_3	17	L	Hi	Me	C_5
9	M	Lo	Lg	C_6	18	L	Hi	Lg	C_6

Tab. 1 Decision fuzzy rules for equivalent capacity estimation [2]

The connection admission control method based on the above equivalent capacity estimation uses two more parameters – information about overload in network y and cell loss probability p_{loss}, that are defined together with membership functions in [2]. Simulations of such proposed method showed that system utilization was increased up to 11% in comparison with conventional CAC methods based on equivalent capacity with comparable quality of service parameters.

Effective bandwidth method with traffic descriptor compression based on neural network [3]. This CAC method uses mean rate and variance of counts of arriving ATM cells in intervals of exponentially growing lengths as the neural network inputs. On the basis of these inputs neural network classifies aggregated cell flow as feasible (i.e. admit new connection) or as inacceptable (i.e. reject new connection). Artificial neural network is trained by data observed from real traffic so there is no need of any requirements on traffic and network models.
Particular inputs into neural network are often dependent each other, so it is possible to transform these inputs into smaller set of inputs, i.e. compress them, without significant loss of performance. The result is faster decision process of CAC method. In [3] was proposed linear compression that can be achieved by addition of another hidden layer with linear activation function into neural network between input layer and initial hidden layer. Simulations showed that three compressed parameters (inputs) offer the same accuracy of decision process as in the case of neural network with uncompressed inputs.

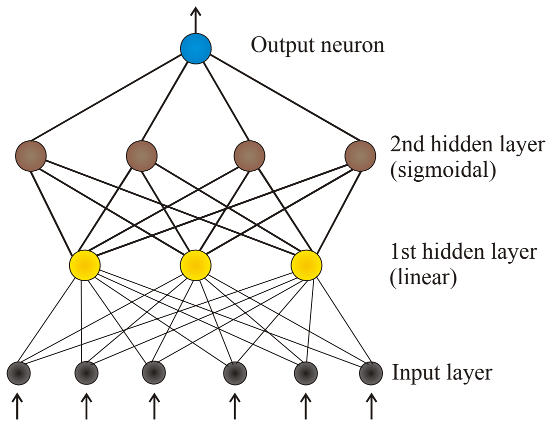

Fig. 3 Neural network with additional hidden linear layer inserted. The new input layer is then the first hidden layer (with three compressed inputs).

The Effective bandwidth method requires only one compressed input parameter *v* that represents effective bandwidth and neural network with one output neuron without hidden layers is assumed. Output of this network is defined as:

$$Y = tgh(v + b_{out}). \tag{13}$$

If *C* is bandwidth available to aggregated flow of connections, then bias of output neuron b_{out} can be regarded as bandwidth *C*. It is assumed that optimal weights α_k will be obtained by learning process and threshold value will be set to *0*. Then new connection is admitted only if $Y \geq 0$, where *Y* denotes output of neural network corresponding to aggregated flow that contains also this new connection. It follows [3] that new connection request is accepted if:

$$v + b_{out} \geq 0 \text{, or } -v \leq b_{out} = C . \tag{14}$$

If we denote *v'* = -*v*, then we have $-v \leq C$, what is in fact decision rule used in the Effective bandwidth method.

4. CONCLUSION

Capabilities of the ATM technology in the area of quality of service and traffic management are undoubtably very sophisticated in comparison with other technologies. Various types of traffic management functions are standardized. However, today, expensiveness of ATM technology is its main disadvantage which inhibit from its wider employment. Nevertheless, the area of traffic management is still evolving and it is subject of many studies and research works. New connection adsmission control methods, especially based on artificial intelligence, are still developing. Hence, it is possible that ATM technology will reappear in the future, but maybe with some modifications.

REFERENCES

[1] Baroňák, I., Kajan, R.: Quality of ATM Services and CAC Methods. Department of Telecommunications, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology Bratislava, Slovak Republic, 1999 (in Slovak).

[2] Cheng, R.G. – Chang, J.: Design of a Fuzzy Traffic Controller for ATM Networks. In: IEEE/ACM Transactions on Networking, vol. 4, no. 3, August 1996, pp. 460-469.

[3] Jordan, M. – Jacobs, R.: Hierarchical Mixtures of Experts and the EM Algorithm. Massachusetts Institute of Technology, Artificial Intelligence Laboratory, A.I. Memo no. 1440, August 6, 1993.

[4] Kouvatsos, D.: Performance Evaluation and Applications of ATM Networks. Boston Kluwer Academic Publishers, 2000, ISBN 0-79-237851-2.

[5] Marzo, J. L. et al.: Comparison of a Convolution and a Fuzzy Logic based CAC. In: IEEE Information Theory Society, 1997, p. 142-149.

[6] Marzo-Lázaro, J. L.: Enhanced Convolution Approach for CAC in ATM Networks, an Analytical Study and Implementation. Doctoral Thesis, Universitat de Girona, Italy, 1996. ISBN 84-8458-103-3.

[7] Shiomoto, K. – Yamanaka, N. – Takahashi, T. : Overview of Measurement-Based Connection Admission Control Methods in ATM Networks. In: IEEE Communications Survey, I/1999, pp. 2 – 13.

[8] Jakab, F.: Pasívne metódy merania prenosových charakteristík sietí pre potreby vyhodnocovania parametrov kvality služieb. Ročenka 2004, Fórum pre komunikačné technológie v Slovenskej republike, str. 83-92, UVTIP Nitra – Vydavateľstvo NOI Bratislava, ISBN 80-89088-36-8.

[9] Hronský, M., Jakab, F., Potocký, M., Jakab, R., Giertl, J.: Sampling Algorithms for Nonintrusive Measurement in Network Oriented Educational Systems, Conference Proceedings of the 4th International Conference on Emerging e-learning Technologies and Applications ICETA 2005, Information and Communications Technologies in Education, Košice, Slovak Republic, 13 - 14 September 2005, Košice, elfa, s.r.o., 2005, 4, pp. 165-171, ISBN 80-8086-016-6.

[10] Jakab, F.: Koncepcia merania a vyhodnoco-vania parametrov QoS, Ročenka 2002, Združenie ATM v SR, Bratislava, UVTIP Nitra - Vydavateľstvo NOI, 2002, pp. 88-102, ISBN 80-89088-11-2.

BIOGRAPHIES

Ivan Baroňák was born in Žilina, Slovakia, on July 1955. He received the electronic engineering degree from the Slovak Technical University Bratislava in 1980. Since 1981 he has been a lecturer at Department of Telecommunications STU Bratislava. In 1992 he submitted PhD work from the field of Terminal telephone equipment. In 1995 he became an associate professor for the subject applied information. Nowadays he works as an associate professor in Department of Telecommunications of FEI STU in Bratislava. Scientifically, professionally and pedagogically he focuses on problems of digital switching systems, ATM, Telecommunication management (TMN), Next Generation Networks, problem of optimal modeling of private telecommunication networks and services.

Matej Kavacký was born in Nitra, Slovakia, in 1979. He received the engineering degree in telecommunications in 2004 from Faculty of Electrical Engineering and Information Technology of Slovak University of Technology (FEI STU) Bratislava. Since 2004 he is PhD. student at the Department of Telecommunications FEI STU and his scientific research is focused on quality of service in broadband telecommunications networks.