A HOP-BY-HOP, ER-BASED ALGORITHM APPLIED IN ABR SWITCHES.

Georgios I. Mousadis
Theodore A.Tsiligirides

Informatics Laboratory, Dept. of General Sciences,
Agricultural University of Athens,
75 Iera Odos, 11855 Athens, Greece

Michael P. Bekakos

Dept. of Informatics,
Athens University of Economics and Business,
76 Patision Street Athens, Greece

Abstract

A new ER-based switch algorithm for flow regulation of the ABR traffic in an ATM network, using a hop by hop strategy, is derived and discussed. Using this scheme, the network determines the load factor and the available capacity in each link. The scheme takes into account both the queue length and the growth rate of the queue length. As a result it explicitly acknowledges the sources about the rates at which they should transmit. The analysis is made possible with the use of a fluid approximation theory. Preliminary numerical results show that the algorithm works properly and that in steady state is stable focus. Therefore, its main purpose, namely to sponge all the available bandwidth for the ABR service is achieved.

Introduction

Asynchronous Transfer Mode (ATM), as a platform refers to a high-bandwidth, low-delay switching and multiplexing technology that can be used in both public and private network applications. It is a fast packet switching method developed originally for the public Broadband Integrated Services Digital Networks (B-ISDN) and it is designed to support all types of traffic streams (voice, video and data), with various characteristics, and different performance requirements. Connections should provide a guarantee Quality of Service (QoS), which is specified mainly in terms of cell delays and cell loss ratio. For new connections the QoS is specified during a connection setup procedure, during which the user must declare its traffic characteristics (at least the peak cell rate). Based on this information, the network rejects the request if the estimated resources are not sufficient. Acceptance of the connection obligates the network to provide the specified QoS and throughput, and the user is obligated to limit its traffic rate accordingly. This agreement forms the basis of the so called traffic contract between the user and the network, which guarantees a specific service scheme that when the user follows shall gain the minimum cell loss. Due to the absence of feedback control during the connection the approach is usually referred as open-loop control. This policy applies very well in services using Constant Bit Rate (CBR) and Variable Bit Rate (VBR). QoS is guaranteed for many traffic parameters, such as the delay, the cell loss ratio, the bandwidth, etc., and it is negotiated at call setup through the admission control and the bandwidth allocation.

However, the open-loop control approach is not appropriate for many data applications. Unlike voice and video, the traffic characteristics of data applications are very difficult to estimate. The Peak Cell Rate (PCR) parameter may be the only predictable parameter, which is not enough information for the network to allocate resources efficiently. In addition, non-real time data applications may adapt to time-varying throughput and tolerate unpredictable cell delays. For these types of applications the Available Bit Rate (ABR) service guarantees a Cell Loss Ratio (CLR) only to those connections in which the source rate dynamically adapts the traffic in accordance with feedback received from the network (closed-loop control). The introduction of ABR service has been motivated by the need of sharing the available bandwidth among all active users, under traffic generated by highly burst data applications. Since in most of the cases data applications cannot predict their own traffic parameters at call setup, an explicit guarantee of service would be wasteful. As a result the network, through the ABR service, provides the *best effort* service scheme, in the sense that no hard QoS guarantees are given, but the network does its best to minimize both the CLR and the delay. A Minimum Cell Rate (MCR) is guaranteed in order to qualify the Virtual Path/Channel Connection (VPC/VCC) as active, provided that the connection is responding to congestion feedback indications from the network. Note, that the ABR traffic has access to bandwidth only when no CBR/VBR traffic is waiting for transmission, namely to bandwidth that would otherwise be unused. Thus the link utilization is increased without affecting the QoS of CBR/VBR connections.

The traffic management group of the ATM Forum had adopted the rate-based approach as a standard for congestion control of the ABR traffic. Rate-based schemes use feedback information from the network to control the rate at which can emit cells into the network. The control information is conveyed to the endpoints through special control cells called Resource Management (RM) cells, which are sent periodically by the sources and turned around by the destinations. The current RM cell format has also been defined by the ATM Forum and includes many fields [2], among which the Explicit Rate Field (ERF), which shows the rate provided by the network at a particular time instant. In addition, the ATM Forum has also specified the source and the destination behaviors, while the behavior at the switch is left to the manufacturers. As a result, the main task of the switches is to determine their load, compute and divide the available bandwidth fairly among active flows and finally to determine the actual Explicit Rate and send it to the source, through the Backward Resource Management (BRM) cell. The current specification is based on the ideas of the ER mechanisms. These are the Enhanced Proportional Rate Control Algorithm (EPRCA) [10, 5], the two congestion avoidance algorithms, known as the Explicit Rate Indication for Congestion Avoidance (ERICA) [8] and the Congestion Avoidance using Proportional Control (CAPC) [3] and the ER based on bandwidth demand estimate algorithm [2].

Apart of ERICA all the above algorithms require setting of many parameters. Incorrect setting may lead to performance degradation. All the above schemes may result in unnecessary oscillations because of the proportional nature of the algorithm used. In addition those schemes using the queue length as overload indicator may lead to unfairness, because the sources that start up late are found to get lower throughput than those which start early. Our interest will be turned into ERICA, which is using few parameters that can be easily tuned [2]. However ERICA cannot always achieve max – min fairness [4], while constant functions restrict the system utilization to a maximum of utilization in the steady state. Thus the system cannot achieve a queuing delay target and it does not provide compensation when measurement and feedback are affected by errors.

The novelty of this work is to propose a scheme for the derivation of the ER field, using functions of both the queuing delay (i.e. the queue length of the switch) and the growth rate of the queue length. Designation of such functions is an interesting study introduced in this work. The proposed ER-based switch algorithm utilizes almost 100% of the required by the ERF bandwidth used in the ABR service. The organization of this paper is as follows. In Section 2, the description and the analysis of the proposed model is presented. The algorithm is solved and analyzed in Section 3. It also includes some numerical results and the available hints for the stability analysis. Finally, the conclusions are presented in Section 4.

MODEL DESCRIPTION AND ANALYSIS

In this section a new ER scheme is described and analyzed in detail. The network model consists of M switches in tandem (Figure 1), allowing the Virtual Source/Virtual Destination (VS/VD) property [2]. This option allows a

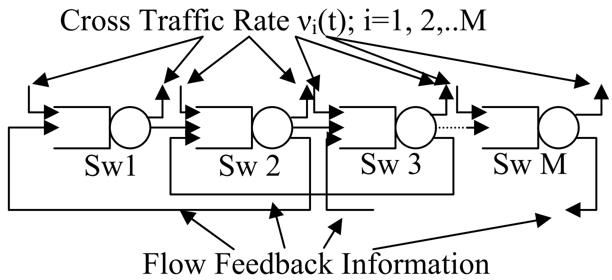

Fig. 1: Model of the Hop-by-Hop strategy (M ATM switches in tandem)

switch to divide an end-to-end ABR connection into separately controlled ABR segments by acting like a destination on one segment, and like a source on the other. Therefore, the network traffic is contributed by Source/ Destination pairs of consecutive switches in the cascade queue and for each such connection there is a VC associated on the path. Due to the above property some hardware complexity may be added.

However, the flow control does not depend on traffic other than that found along the path of the VC. This simplifying assumption means that the cross traffic entering a switch does not proceed through the tandem queues but it is renewed at each stage, i.e., it immediately leaves the system. Particularly, in this model, it is assumed that a switch may serve two types of ABR traffic streams; a tagged traffic having the highest priority and a background traffic, which immediately leaves the switch. Every VS acknowledges the ERF of the corresponding BRM cell and adjusts its transmission rate every round trip time by calculating the Mean Arrival Cell Rate (MACR) for the time interval that follows.

To clarify the notation we assume that there are L links with each link i characterized by a transmission capacity $c_i = 1/t_i$ (cells/sec), a propagation delay τ_i and a processing capacity $1/tpr_i$ (cell/sec), where tpr_i is the internal delay of the switch i, namely the time needed for a cell to be proceeded from the input to the output queue. It is asumed that the processing capacity of each node is much larger than the total transmission capacity of its incoming links. Thus, the only reason causing congestion is the transmission capacity.

Each link maintains a separate queue for each VC passing through it. It is assumed that the occupancy at time t of the queue associated with link i and $VC\ j$ is $X_{i,j}(t)$ with $xmax_{i,j}$ the corresponding queue threshold level. Each source is assumed to transmit at its maximum transmission speed $\lambda_s = 1/t_s$ and the control law computes the source input rate $\lambda(t)$(cells/sec). The bandwidth delay product τ_i / t_i, represents the number of *in flight* cells on the transmission link.

To model the dynamic behavior of each queue it is assumed a deterministic fluid approximation of cell flow.

To simplify the analysis, which regulates the source rates only one VC is considered and therefore the subscript j may be dropped. To control the queue length $X_i\,(t)$ for a specific VC, a simple controller is used. The state of the controlled connection at the corresponding switch i is fully captured by three state parameters, namely the Mean Arrival Cell Rate $MACR_i$, the Allowed Cell Rate ACR_i and the buffer occupancy X_i . Thus, the ERF_i of the corresponding BRM cell produced by the i^{th} switch, may be written as

$$ERF_i(t_n) = F\{X_i(t_n), MACR_i(t_n), ACR_i(t_n)\} \qquad (1)$$

where, $ERF_i(t_n)$, $ACR_i(t_n)$ and $X_i(t_n)$ denote the corresponding parameters at the time $t_n = nT$, $n = 0, 1, 2$, .. and T is the sampling time, while $MACR_i(t_n)$ denotes the mean arrival cell rate during the period $[(n\text{-}1)T, nT)$. The ATM Forum [2] suggested:

$$ERF_i(t_n) = \min\{ERF_i(t_n), ERF_{i+1}(t_n)\}\tau_1 \qquad (2)$$

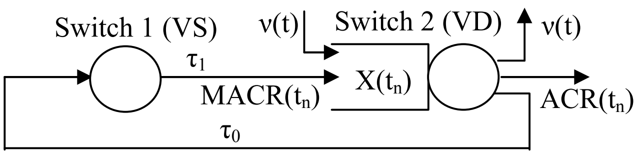

Fig. 2: The single Hop single VC model. Here, the background traffic has a priority rate $\nu(t_n)$. Fixed Round Trip time FRT = $\tau = \tau_0 + \tau_1$.

and that in such systems $ACR_i(t_n)=MACR_{i+1}(t_n)$ for all i. Without loss of generality the analysis may restricted to the single hop single VC model (see Fig. 2). Thus, the index i may be dropped. The switch is assumed to have a large but finite buffer to use, of volume $Xmax$. The Fixed Round Trip (FRT) time $\tau(= \tau_1 + \tau_0)$ is the propagation delay experienced by the cells before they reach the bottleneck queue of the VD, say τ_1, plus the propagation delay experienced by the BRM cells before returning to the VS, say τ_0. Further, $ACR(t_n)=\mu\text{-}\nu(t_n)$, where μ denotes the fixed service rate of the switch and $\nu(t_n)$ is the exogenous traffic rate at time t_n which is not constant. In general, $ACR(t_n)$ is not a linear function of the parameters $MACR(t_n)$ and $x(t_n)$. The new rate is calculated at the switch and it is provided to the VS through the ERF of the BRM cells, which are delivered every FRT period. The VS is always able to adjust its transmission rate to the required level through the function:

$$MACR(t_n + \tau_1) = \\ = \min\{MAR, \min\{PCR, \max\{MCR,\ ERF(t_n - \tau_0)\}\}\}, \qquad (3)$$

where MAR is the maximum allowed link rate. Notable, that in the worst case of traffic:

$$ERF(t_n - \tau_0) = MACR(t_n + \tau_1), \text{ or}$$

$$ERF(t_n) = MACR(t_n + \tau) \qquad (4)$$

NUMERICAL ANALYSIS

Let $K(t_n)$ be the amount of the outstanding unacknowledged cells in the VS, at the time $t_n = n\tau$, n=0, 1, 2.. . In fact, $K(t_n)$ presents the number of cells forwarded to the switch during the $[(n\text{-}1)\tau, n\tau)$ period plus the number of cells already waiting in the buffer. Thus, in case the observer is located at the VS site the system is described by the following equation:

$$K(t_n) = \tau\, MACR(t_{n-1}+\tau_1) + X(t_n\text{-}\tau_0) \qquad (5)$$

Differentiating with respect to t, one may obtain:

$$\frac{d}{dt}K(t_n) = \tau\frac{d}{dt}MACR(t_{n-1}+\tau_1) + \frac{d}{dt}X(t_n - \tau_0) =$$

$$= \tau\frac{d}{dt}MACR(t_n - \tau_0) + \frac{d}{dt}X(t_n - \tau_0)$$

Assuming $\Delta K(t_n) = A\ X(t_n\text{-}\tau_0) + B$, and taking into account that the time between adaptation is $\Delta t = \dfrac{u}{ACR(t_n - \tau_0)}$, it follows:

$$\frac{d}{dt}K(t_n) \approx \frac{\Delta K(t_n)}{\Delta t} = [AX(t_n - \tau_0) + B]\frac{ACR(t_n - \tau_0)}{u}.$$

In the above, B represents the internal processing delay of the switch and u is the number of BRM cells every FRT period. Therefore, in case $A \neq 0$, we obtain:

$$\frac{d}{dt}MACR(t_n - \tau_0) = \left[X(t_n - \tau_0) + \frac{B}{A}\right]\frac{A}{u\tau}ACR(t_n - \tau_0) \\ -\frac{d}{dt}\frac{X(t_n - \tau_0)}{\tau} \qquad (6)$$

Similarly, assuming that the observer is located in the VD site and using the equation (5), the system may be described as follows:

$$\frac{d}{dt}MACR(t_n) = \left[X(t_n) + \frac{B}{A}\right]\frac{A}{u\tau}ACR(t_n) - \frac{d}{dt}\frac{X(t_n)}{\tau} \qquad (7)$$

Thus, the non-linear ODEs system controlling the transmission rate of the VS and the buffer occupancy at the switch is given by:

$$\frac{d}{dt}ERF(t_n) = \begin{cases} \dfrac{B}{u\tau}ACR(t_n), & \text{if } (a) \\ \dfrac{B}{u\tau}ACR(t_n) - \dfrac{d}{dt}\dfrac{X(t_n)}{\tau}, & \text{if } (b) \\ \left[X(t_n) + \dfrac{B}{A}\right]\dfrac{A}{u\tau}ACR(t_n) - \dfrac{d}{dt}\dfrac{X(t_n)}{\tau}, & \text{if } (c) \\ \left[X(t_n) + \dfrac{B}{A}\right]\dfrac{A}{u\tau}ACR(t_n), & \text{if } (d) \end{cases} \qquad (8)$$

$$\frac{d}{dt}X(t_n) = \begin{cases} MACR(t_n) - ACR(t_n), & \text{if (e)} \\ 0, & \text{otherwise} \end{cases} \qquad (9)$$

In the above (a) - (e) are given as follows:

$(a):\quad X(t_n) = 0 \wedge MACR(t_n) < ACR(t_n).$

$(b):\quad X(t_n) = 0 \wedge MACR(t_n) \geq ACR(t_n).$

$$(c):\ \begin{cases} 0 < X(t_n) < X\max, \text{or} \\ X(t_n) = X\max \wedge MACR(t_n) \le ACR(t_n) \end{cases}.$$

$$(d):\ \text{is}\, X(t_n) = X\max \wedge MACR(t_n) > ACR(t_n).$$

$$(e):\ \begin{cases} 0 < X(t_n) < X\max\ \text{or}, \\ X(t_n) = 0 \wedge MACR(t_n) \ge ACR(t_n)\,\text{or}, \\ X(t_n) < X\max \wedge MACR(t_n) \le ACR(t_n) \end{cases}.$$

Assuming that the $ACR(t_n)$ is constant during the time period $[n\tau, (n+1)\tau)$, $\forall\ n \in N$, the system of ODEs described by the equations (8) and (9) may be rewritten as follows:

$$\frac{d}{dt}MACR(t_n) = \begin{cases} X_{st}, & \text{if } (a) \\ X_{st} - \frac{d}{dt}\frac{X(t_n)}{\tau}, & \text{if } (b) \\ [X_{st} - X(t_n)] - \frac{d}{dt}\frac{X(t_n)}{\tau}, & \text{if } (c) \\ [X_{st} - X(t_n)], & \text{if } (d) \end{cases},$$

$$\frac{d}{dt}X(t_n) = \begin{cases} MACR(t_n) - ACR(t_n), & \text{if } (e) \\ 0, & \text{otherwise} \end{cases}$$

where, $X_{st} = \frac{B}{u\tau}ACR(t_n)$.

The above system is solved numerically using the Euler predictor method. The algorithm is as follows:

$$MACR(t_{n+1}) = \begin{cases} MACR(t_n) + X_{st}\tau, & \text{if } (a) \\ X_{st}\tau + ACR(t_n), & \text{if } (b) \\ [X_{st} - Xpr]\tau + ACR(t_n), & \text{if } (c) \\ MACR(t_n) + [X_{st} - Xpr]\tau, & \text{if } (d) \end{cases},$$

Where, Xpr is given by:

$$Xpr = \begin{cases} X(t_n) + (MACR(t_n) - ACR(t_n))\tau, & \text{if } (e) \\ X(t_n), & \text{otherwise} \end{cases}$$

Note that Xpr is calculated at time t_n and represents the prediction of the buffer at time$(t_n+\tau)$.

As it appears, starting with an initial buffer $X(t_0)$ and $MACR(t_0)$ and taking into account the equation (4), the procedure derives the $ERF(t_0)$ using the Euler predictor method. This value is used to calculate the new buffer $X(t_1)$ (presented in the algorithm as Xpr), which is then used to produce the new $ERF(t_1)$ and so on. Note that the switch sends u BRM cells every time period τ with the same ERF value. This is because the corresponding $ACR(t_n)$ is assumed constant during the time period [nτ, (n+1)τ). Further, in the steady state the buffer occupancy is given by:

$$X(t_{n+1}) = \begin{cases} X(t_n) + (MACR(t_n) - ACR(t_n))\tau, & \text{if } (e) \\ X(t_n), & \text{otherwise} \end{cases}$$

Finally, assuming as the worst case of traffic the situation when the buffer overflows, say with rate $BO(t_n)$, we have:

$$\frac{d}{dt}BO(t_n) = \begin{cases} MACR(t_n) - ACR(t_n), & \text{if } (f) \\ 0, & \text{otherwise} \end{cases}, \qquad (10)$$

where,

$$(f):\ X(t_n) = X\max \wedge MACR(t_n) \ge ACR(t_n)$$

Numerical results are given in Figs. 3(a) and 3(b), where

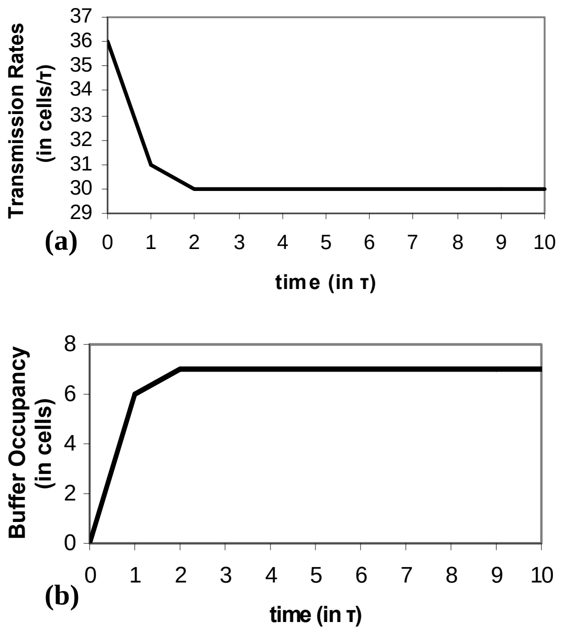

Fig. 3: The model with ACR(t_n) constant (i.e. steady state). Link Rate = 155,52 Mbps. ACR(t_n) =30 cells/τ, PCR = 36 cells/τ, MCR = 1 cell/τ and τ=100 μsec. Initial Condition (MACR(0), X(0))=(PCR, 0).

the transmission rate ($MACR$) and the buffer occupancy versus the time are presented. From the above analysis one may point out that the linear branch of the system (8) and (9) has a stable focus point (the relevant theory may be seen in [1] pg. 135), provided the $ACR(t_n)$ remains constant for some time period and the bandwidth overflow rate in equation (10) is zero (steady state conditions). The results also verify that the system has a stable focus point given by:

$(MACR, X) = (ACR(t_n), X_{st}) = (ACR(t_n), -B/A)$,

where the optimal value for the parameter A is given as $A = -(u\,\tau/ACR(t_n))$.

CONCLUSIONS

The study of the ER-based feedback schemes has revealed that the queuing delay and the growth rate of the queue length are working together to effect the calculation of the ERF. In this sense both factors need to be taken into consideration in the design algorithm used to regulate the best effort traffic (ABR service) in an ATM network.

In this work we developed a new ER-based switch algorithm that takes into account the above factors. As a result it utilizes almost 100% of the bandwidth allocated for the ABR service. It was shown that the proposed algorithm dynamically adjusts the adaptive VS window size by regulating the emission rate, through the ERF of the BRM cells. The model was analysed through a non-linear system of ODEs, assuming that the ACR, MACR, and the VD's buffer occupancy were modeled as fluids. It was also appeared that in steady-state those systems are stable focus. A simple numerical method by means of the Euler predictor method for solving such a system was derived. It appeared that using the proposed ER-based switch algorithm, the oscillating behavior of the system was eliminated.

Work is still in progress to provide a comparison of the above scheme with the ERICA, the most popular algorithm used so far, and to examine fairness behavior.

REFERENCES

1. Arnold V. I., Ordinary Differential Equations, MIT Press edition, 1978 ISBN 0-262-51018-9.
2. ATM FORUM. The ATM Forum Traffic Management Specification Version 4.0 ftp://ftp.atmforum.com /pub/approved-specs/af-tm-0056.000.ps, April 1996
3. Barnhart A.W., 'Explicit Rate Performance Evaluation', ATM Forum AF94-0983R1 http://www.cis.ohio-state.edu/~jain/atmforum.htm
4. Danny Tsang H. K., Wales Kin Fai Wong, Sheng Ming Jiang and Eric Y. S. Liu, 'A Fast Switch Algorithm for ABR Traffic to Achieve Max-Min Fairness', International Zurich Seminar on Digital Communications, IEEE, February 19-23, 1996.
5. Hsiaw H., Roberts S., Jain A., Makrucki B., Tofigh T., Barnhart A.W., Holden B., Fedorlow G., Daigle J.N., Kline D., Suzuki H., Ramamurthy G., Giroux N., Kositpaiboon R., Hughes D., Hluchyj M.G., Garg G.,Siu K.Y., Tzeng H.Y. and Yin N., 'Closed-Loop Rate-Based Traffic Management', ATM Forum Contribution 94-0438R2, September 1994.
6. Jain R., Kalyanaraman S., Fahmy S., Goyal R., 'Source Behaviour for ATM ABR. Traffic Management: An Explanation', IEEE Communications Magagine, November 1996.
7. Jain R., 'Congestion Control and Traffic Management in ATM Networks: Recent Advances and a Survey', Computer Networks and ISDN Systems, October 1996
8. Jain R., Kalyanaraman S, Viswanathan R., Goyal R., 'A Sample Switch Algorithm', ATM Forum AF95-0178R1, February 1995, http://www.cis.ohio-state.edu/~jain/atmforum.htm
9. Jain R., Kalyanaraman S. and Viswanathan R., 'Rate Based Schemes: Mistakes to Avoid', ATM Forum/94-0882, September 1994, http://www.cis.ohio-state.edu /~jain/atmforum.htm
10. Roberts L., 'Enhanced PRCA', ATM Forum Contribution 94-0735R1, July 1994.