An Adaptive Polling Scheme for Multiple Access to Wireless ATM Networks

Sung Won Kim[1], Dong Geun Jeong[2], and Chong-Ho Choi[3]

[1]School of Electrical Eng., Seoul National Univ., Seoul, 151-742, Korea (email: ksw@csl.snu.ac.kr), and Media & Communication Research Center, LG Electronics, Korea

[2]Dept. of Electronics Eng., Hankuk Univ. of Foreign Studies, 449-791, Korea (email: dgjeong@san.hufs.ac.kr)

[3]School of Electrical Eng., Seoul National Univ., Seoul, 151-742, Korea (email: chchoi@csl.snu.ac.kr)

Abstract— **This paper proposes a polling scheme for multiple access to wireless ATM networks. By combining the concepts of virtual source and adaptive polling, the proposed scheme serves effectively the constant bit rate (CBR), the variable bit rate (VBR), and the available bit rate (ABR) traffics. To maximize the (radio) channel efficiency, the proposed scheme is designed to guarantee the stochastic delay bound based on the concept of equivalent capacity. The relation between the supplied capacity, design parameters, and cell loss probability is analyzed and compared with simulation results. As a results, it is shown that the proposed scheme is so simple and efficient to be used in wireless ATM environments.**

I. Introduction

The asynchronous transfer mode (ATM) has become one of leading technologies for future wireless multimedia communications. Specifically, the "ATM air-interface" as well as the ATM (fixed) core network is essential for "fully ATM-based wireless networks."

In a wireless ATM system, a cell consists of a base station (BS) and several wireless terminals. The BS functions both as a centralized controller for the cell and as a gateway to the fixed network. Our main concern in this paper is the multiple access of the uplink (from terminals to BS) at the air-interface since the BS is an unique transmitter on the downlink (from BS to terminals) and can appropriately make out a schedule for its transmission.

From the viewpoint of multiple access scheme for wireless networks, the code-division multiple access (CDMA) is appropriate to provide high system-wide capacity using scarce (radio) resource. However, the time-division multiple access (TDMA) is more desirable for wireless ATM since it is more flexible in resource allocation [1], which is indispensable for managing the multimedia traffic with high date rate. Thus, some proposals have suggested CDMA with time-division duplex (CDMA/TDD) scheme, having the merits of both CDMA and TDMA [2]. Although the concepts presented in this paper can be applied to any of CDMA, TDMA, and CDMA/TDD, we consider hereafter only TDMA systems for convenient description.

The multiple access techniques to time-divided channel can be classified into the fixed assignment, the random access, and the demand assignment. The demand assignment schemes can be categorized again into the reservation methods and the polling methods [3]. With the centralized architecture such as in wireless ATM, the polling methods are simpler than the reservation methods since there is no reservation phase. However, the performance of polling methods depend heavily on the algorithm that manages the polling period for each connection. Based on whether the polling period is fixed or variable, we can subdivide the polling methods into "fixed" and "adaptive" methods.

The wireless ATM networks should support various classes of traffic, i.e., the constant bit rate (CBR), the variable bit rate (VBR), the available bit rate (ABR), and the unspecified bit rate (UBR) traffics [4]. Moreover, some of these traffics require the dynamic data rates. In this situation, adaptive polling methods are preferred.

In this paper, we propose a new adaptive polling scheme for wireless ATM system. In Section II, we propose the polling scheme to bound the cell delay. The cell loss probability with the proposed scheme is analyzed and compared with the simulation results in Section III. Finally, concluding remarks are offered in Section IV.

II. Polling Mechanism Design

We use the *virtual traffic source* (VS) at the BS, which emulates the corresponding *real source* (RS) at terminal [5], [6]. A VS generates tokens whereas a RS generates packets. The VS's, instead of RS's, request the bandwidth by the tokens. One token requests the polling of one packet in corresponding connection by one ATM cell. Thus, the size of one packet is equal to the payload size of one ATM cell. The packet (token) queue buffers the packets (tokens) until they are retrieved. The packets (tokens) in queue are served by first-in-first-out (FIFO) manner. At a connection setup, the terminal makes RS's packet queue and the BS sets up a VS and its token queue (buffer) that correspond to the requesting connection.

A. Virtual Source Design

We first suggest a VS model that can be applied to any of CBR, VBR, and ABR. This model is based on "two-states rate process."

For the convenient explanation of the model, let us consider the simplest rt-VBR traffic such as voice traffic.

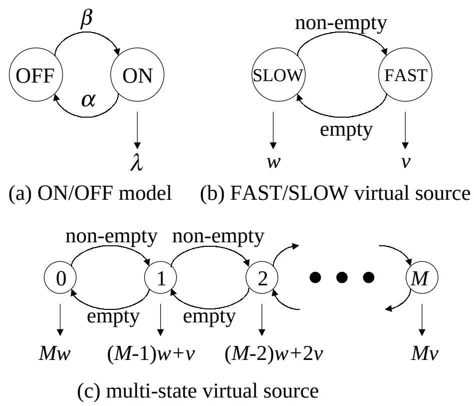

Fig. 1. Virtual source model.

The RS of voice traffic can be modeled as an ON/OFF source [Fig. 1(a)]. The source generates packets at a constant rate λ (in packets/sec) while it is in ON state. In OFF state, it does not generate packet. We assume in the followings that the ON and OFF periods are exponentially distributed, respectively. Let α and β respectively denote the state transition rates from ON to OFF and from OFF to ON.

The VS has two states similar to the ON/OFF RS model. We call the two states of VS, "FAST" and "SLOW" [Fig. 1(b)]. Tokens are generated at a constant rate v (in tokens/sec) in FAST state, and at a constant rate w (in tokens/sec) in SLOW state ($v \geq w$). The state transition occurs by the polling result. When the polled time slot is empty (i.e., the terminal with polled connection does not send valid data) and the current state is FAST, the state changes to SLOW. When the polled time slot is not empty and the current state is SLOW, the state changes to FAST.

Now, we consider more complicated rt-VBR such as the compressed video. There are several traffic models for the compressed video data. One of them is the aggregate of M identical ON/OFF mini sources [7]. Similarly, we propose the multi-state VS that is multiplexed with M homogeneous FAST/SLOW mini VS's depicted in Fig. 1(c). Note that the basic FAST/SLOW VS corresponds to $M = 1$. Each state represents the number of mini sources that are in FAST (sub)state. The state transition occurs by the result of its polling as explained in FAST/SLOW VS. For example, let us consider the transition from state k to $k+1$. In state k, if there is a packet sent in response to the token generated by a mini source in SLOW (sub)state, the mini source enters the FAST (sub)state and the source-wide state changes to the state $k+1$.

We model the VS's for CBR, VBR, and ABR, by the above FAST/SLOW VS with the proper setting of parameters:

- For CBR, $v = w$ and the values are equal to the peak cell rate (PCR) of connection.
- For VBR, v is set to PCR/M. The decision of w depends on QoS requirement. The effect of w will be discussed in the next section.
- For ABR, w is set to the minimum cell rate (MCR), and v can increase or decrease by resource management (RM) cells that are used for feedback flow control in ATM networks [4].

At call setup phase, the BS decides VS model and the values of v, w, and M, based on the traffic characteristics. The appropriate value of M depends highly on the characteristic of RS. Thus, to deal with M accurately, some extensive considerations on the RS (e.g., video traffic) itself are needed. On the other hand, the feedback flow control of ABR is not the focus of this paper. Thus, to concentrate our attention on the polling scheme and to simplify the analysis, we hereafter consider only the basic FAST/SLOW VS ($M = 1$) and the corresponding ON/OFF RS of CBR and VBR traffics. Hence, we will set the value of v as

$$v = \lambda. \tag{1}$$

B. Packet Queue of Real Sources and Delay Bound

For real time traffics such as CBR and rt-VBR, packets delivered with too long delay lose their values. The delivery of these packets wastes the scarce bandwidth of wireless network. Thus, any packet held beyond a certain delay bound, D_{max}, is discarded by the terminal. The discarded packets are considered to be lost.

To implement the packet discard mechanism easily, we propose the following algorithms.

- Each packet header contains the time stamp of its arrival time to packet queue.
- Only when the packet is retrieved for polling, its arrival time is checked. If the time difference between the arrival time and the current time is more than D_{max}, it is discarded and the next packet in the queue is retrieved. We call this method "check-out."
- When the newly arriving packet finds the queue full, the oldest packet is discarded and the arriving packet is queued. We call this method "push-out."

Now, we assume that

(A1) the duration of ON state is longer than the interarrival time of tokens generated at SLOW state ($1/w$) and

(A2) the tokens generated at SLOW state are not lost.

Note that, with appropriate values of w, (A1) is acceptable in practical situations. The condition for (A2) will be discussed in next subsection. Under these assumptions, when the packet queue is not empty, the state of RS is ON and/or the state of VS is FAST. In other words, when the queue is not empty, the queue is being filled with packets at rate λ and/or the packets are being retrieved at rate v. Thus, for a given queue size, we can bound the cell delay only with push-out. In this case, to bound the cell delay within D_{max}, the packet queue size, Q_p, should be set to as

$$Q_p = \lfloor \lambda D_{max} \rfloor = \lfloor v D_{max} \rfloor. \tag{2}$$

The second equality comes from (1). From now on, we implement the packet queue such that the packet queue size is set as (2) and packets are discarded by push-out. This implementation method eliminates the terminal's overhead of real time processing for check-out and time stamp.

For non-real time traffics, the packet queue size can be appropriately large since there is no bound on cell delay.

C. Clustered Token Queue and Delay Bound

The connections belonging to the same traffic class and having the same QoS requirements can be grouped into a cluster. The VS's belonging to a cluster share *one* token queue. Thus, in the description of the proposed scheme, we assume that one QoS cluster has just one token queue. Let C_k (in cells/sec) be the supplied capacity to the cluster k ($1 \leq k \leq K$) where K is the number of clusters. We use the unit of capacity tokens/sec and packets/sec interchangeably in this paper. But cells/sec has different meaning because an ATM cell is composed of a packet and overhead. Thus we define the ratio, r, as payload size divided by ATM cell size. Then, the amount of capacity that is used for packet transmission (or, equivalently, token retrieval) in C_k is defined as

$$\hat{C}_k = rC_k \tag{3}$$

and the unit of $\hat{C}_k$ is packets/sec or tokens/sec. The capacity of $(1-r)C_k$ is used for overhead transmission.

We assume that $N_k w \leq \hat{C}_k$ where N_k is the number of VS's that belong to cluster k. Like the packet queue, the clustered token queue also uses the push-out to bound the tokens' delay. Then, the clustered queue size, Q_c, is given as

$$\begin{aligned} Q_c &= \lfloor \hat{C}_k D_{max} \rfloor \\ &= \lfloor rC_k D_{max} \rfloor. \end{aligned} \tag{4}$$

It is noted that although using the clustered token queue structure, we can not cluster the packet queues in several terminals and each packet queue size is determined by (2). Hereafter, the subscript k is used for the indication of the cluster k.

III. Cell Loss Probability

In the proposed scheme, the most cell loss can occur in two cases.

- If there is sufficient capacity, token loss cannot occur. However, even though this is the case, the BS may poll terminal too slow at SLOW state and, thus, the packet queue in the RS can be overflowed during the time period between the RS state transition from OFF to ON and the VS state transition from SLOW to FAST. In this situation, packets may be lost by push-out.
- If the polling rate at SLOW state (w) is fast enough, there is no packet loss during the time period mentioned above. However, although this is the case, if the supplied capacity is not sufficient, tokens may be lost by push-out, which result in the packet loss.

We analyze these two effects separately. First, we calculate the cell loss probability (CLP) of the proposed scheme for a given polling rate under the assumption of sufficient capacity. Second, we estimate CLP for a given capacity under the assumption that one token loss incurs only one packet loss. Finally, we merge these two loss probabilities to get the overall CLP for given parameters. For the sake of convenience, it is assumed that the communication channel is ideal.

A. Effect of Token Generation Rate on CLP

We assume the sufficient capacity is given to clustered token queue k so that there is no queueing delay. We assume in addition:

(A3) The duration of OFF state is longer than Q_p/v.

(A4) The packet arrival time and the token generation time are independent to each other.

Note that Q_p/v is the required time to empty (fill) the packet queue that is full (empty) at FAST state of VS and OFF state of RS (at SLOW state of VS and ON state of RS). Assumptions (A3) and (A4) are acceptable practically.

Let the start of OFF state begin at epoch 0. Then, the interarrival time of OFF states are independent and identically distributed nonnegative random variables. Let X_n (Y_n) be the number of packets arrived (lost) between the arrival time of nth OFF state and $(n+1)$th OFF state. Then, $\{X_n, n = 1, 2, \cdots\}$ ($\{Y_n, n = 1, 2, \cdots\}$) is a sequence of nonnegative independent random variables. Consequently, the time duration of ON/OFF model can be considered as renewal process.

We calculate CLP with a given polling rate at SLOW state, $L_c(w)$, that the packets in terminal experience.

If we set the polling rate at SLOW state as

$$w \geq \frac{v}{Q_p}, \tag{5}$$

then from (2) it is noted that the polling interval at SLOW state is not greater than D_{max}, i.e.,

$$1/w \leq D_{max}. \tag{6}$$

Thus, the number of waiting packets does not exceed Q_p, i.e., there is no push-out. Therefore, $L_c(w) = 0$. We call this operation region the "over-polling" because w is fast enough for zero CLP.

On the other hand, when $w < v/Q_p$ ("under-polling"), we have CLP as

$$\begin{aligned} L_c(w) &= \frac{E(Y_n)}{E(X_n)} \\ &= \Pr(Y_n > 0)\frac{E(Y_n \mid Y_n > 0)}{E(X_n)} \end{aligned} \tag{7}$$

where the first equality follows the property of renewal process.

During the duration of Q_p/v, the queue can buffer the arriving packets without loss. Thus, we can get,

$$\Pr(Y_n > 0) = \frac{w^{-1} - Q_p v^{-1}}{w^{-1}}. \quad (8)$$

From (A4) and the fact that packets are generated only in ON state, we also have

$$\begin{aligned} E(Y_n \mid Y_n > 0) &= v\frac{w^{-1} - Q_p v^{-1} + v^{-1}}{2} \\ &= \frac{v - Q_p w + w}{2w}, \quad (9) \\ E(X_n) &= \frac{v}{\alpha}. \quad (10) \end{aligned}$$

Substituting (8)–(10) to (7), we have

$$L_c(w) = \frac{\alpha(\hat{w} - Q_p)(\hat{w} - Q_p + 1)}{2\hat{w}v} \quad (11)$$

where $\hat{w} = v/w$.

Substituting (2) to (11), we get another form of $L_c(w)$,

$$L_c(w) = \frac{\alpha(\hat{w} - \lfloor vD_{max} \rfloor)(\hat{w} - \lfloor vD_{max} \rfloor + 1)}{2\hat{w}v}. \quad (12)$$

B. Effect of Limited Capacity on CLP

The analysis in the previous subsection is very accurate when the given capacity is sufficiently large. However, the analysis cannot be used when the capacity is not sufficient. In that case, cell loss occurs mainly by the lack of capacity, irrespective of polling region (i.e., over-polling or under-polling). To estimate CLP in this situation, we take another analysis approach that removes the approximation of "sufficient capacity" although it gives more rough results than the previous approach.

Note that the system bandwidth is consumed by tokens that are generated from VS. Thus, we first estimate the token loss probability with limited capacity. Then, we approximate it to CLP based on the assumption that

(A5) one token loss incurs one cell loss.

We can categorize the effect of this assumption as follows.

- When the state of VS is FAST and the queue is full, the assumption is appropriate.
- When the state of VS is FAST and the queue is not full, the token loss incurs no cell loss. The assumption becomes conservative.
- When the packet buffer is not empty, the lost tokens generated at SLOW state may incur more than one cell loss. We do not consider this situation according to (A2).
- When the packet buffer is empty, the token loss incurs no cell loss. The assumption becomes conservative.

Overall, the assumption (A5) is somewhat conservative.

To get the token loss probability of the clustered token queue structure, we assume the token flow is like a continuous flow of fluid. This approximation method is generally used in the calculation of "equivalent capacity" [7], [8]. The concept of equivalent capacity is bandwidth requirement of both individual and multiplexed connections, based on their statistical characteristics and the desired QoS. We follow the works in [7] and [9], which use multiplexed multiple two-state (ON/OFF) fluid-flow sources. However, since both FAST and SLOW states in our VS generate tokens, some modifications should be made.

Let $F_i(t, Q_c)$ denote the probability that i VS's are in FAST state and the queue occupancy (length) does not exceed Q_c at time t. If we approximate that state transition rates of VS's are equal to those of ON/OFF RS's, then we can get the following equation

$$\mathbf{D}\frac{d}{dQ_c}\mathbf{F}(Q_c) = \mathbf{M}\mathbf{F}(Q_c), \qquad Q_c \geq 0 \quad (13)$$

where

$$\begin{aligned} \mathbf{D} &= \mathrm{diag}\{N_k w - C_k, v + (N_k - 1)w - C_k, \ldots, N_k v - C_k\}, \\ \mathbf{M} &= \begin{bmatrix} -N_k\beta & \alpha & & & \\ N_k\beta & -(N_k-1)\beta - \alpha & & & \\ & (N_k-1)\beta & & & \\ & & \ddots & & \\ & & & \beta & -N_k\alpha \end{bmatrix}, \\ \mathbf{F}(Q_c) &= [F_0(Q_c), F_1(Q_c), \ldots, F_{N_k}(Q_c)]. \end{aligned}$$

Then the distribution of the buffer contents is of the form

$$\mathbf{F}(Q_c) = \sum_{i=0}^{N_k} a_i \mathbf{\Phi}_i e^{z_i Q_c} \quad (14)$$

where the z_i and $\mathbf{\Phi}_i$ are, respectively, generalized eigenvalues and eigenvectors associated with the solution of the differential equation (13). The a_i's are coefficients determined from boundary conditions (see [9] for details).

Then the survivor function $G(Q_c)$ that the token queue occupancy exceeds Q_c is

$$G(Q_c) = 1 - \sum_{i=0}^{N_k} F_i(Q_c). \quad (15)$$

Even if $G(Q_c)$ is the tail of the token queue length distribution, we approximate that the token loss probability is $G(Q_c)$. Thus, from (4) and (A5), CLP with limited capacity, $L_t(C_k)$, is given as

$$\begin{aligned} L_t(C_k) &= G(C_k) \\ &= 1 - \sum_{i=0}^{N_k} F_i(\lfloor rC_k D_{max} \rfloor). \quad (16) \end{aligned}$$

C. Overall Cell Loss Probability

Finally, we combine the CLP with limited capacity given by (16) and the previous result of CLP with sufficient capacity (12) to get the overall CLP for given parameters, $L(w, C_k)$. Note that CLP caused by both

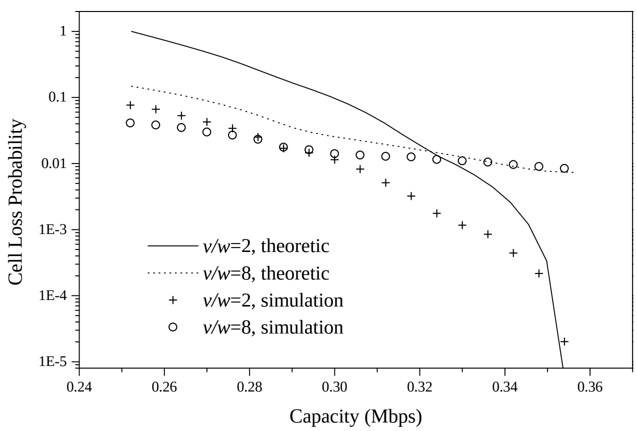

Fig. 2. Overall cell loss probability for given parameters.

limited capacity and slow token generation rate, i.e., CLP caused by lost tokens generated at SLOW state, is not considered in this paper as discussed in previous subsection. Thus, $L(w, C_k)$ is the sum of $L_c(w)$ in (12) and $L_t(C_k)$ in (16), i.e.,

$$L(w, C_k) = L_c(w) + L_t(C_k) \tag{17}$$

under the assumption that the other factors are negligible compared with these two CLP's.

Fig. 2 shows CLP versus supplied capacity given by (17). It was assumed that ATM payload size is 47 bytes, ATM adaptation layer (AAL) overhead is 1 byte, and ATM cell size is 53 bytes. Thus r is 47/53. The nominal values of parameters are: $v = 170$ packets/sec ($= 170 \times 47 \times 8 = 64$ kbps), $\alpha = 1$, $\beta = 2/3$, $N_k = 5$, and $D_{max} = 23.53$ msec. Thus, $Q_p = 4$ by (2). The CLP's, calculated by (17), are depicted in solid or dotted lines and simulation results are also presented.

In over-polling condition ($v/w = 2$), $L_c(w)$ is zero, and $L_t(C_k)$ is equal to $L(w, C_k)$. On the other hand, $L_c(w)$ has nonzero values in under-polling condition ($v/w = 8$), and $L(w, C_k)$ has the limiting value of $L_c(w)$ as the capacity goes to peak rate capacity. As shown in the figure, the difference between the theoretical and simulation results is larger when the capacity is not sufficient. This is because that the assumption (A5) is conservative, as mentioned in previous subsection.

Note that the CLP of under-polling and over-polling are different and crossed. Thus, we can select the optimal value of w that minimizes the capacity requirement depending on the CLP requirement.

IV. Conclusions

We have proposed a new adaptive polling scheme for wireless ATM air-interface. The unified concept of FAST/SLOW VS was applied to all the ATM traffic classes. Since the stochastic delay bound is used, the ATM systems with the proposed scheme exploit the radio resource very efficiently.

The other merits of the proposed scheme can be found in its implementation. First, the scheme is very simple. Second, almost all protocol overhead can be implemented in BS, while the minimum overhead is required in terminals. This is a remarkable virtue for wireless systems, since low production-cost and low power-consumption for wireless terminals can be achieved.

In the analysis, we have assumed only the basic ON/OFF RS (and FAST/SLOW VS) and ideal physical channel to demonstrate the efficiency of the proposed scheme simply. Thus, the further studies related on this paper can include the analysis of multi-state VS and the performance evaluation of the proposed scheme on the erroneous channel. We will undertake some of these works.

References

[1] W. S. Jeon, D. G. Jeong, and C. -H. Choi, "An integrated services MAC protocol for local wireless communications," *IEEE Trans. Veh. Technol.*, vol. 47, no. 1, pp. 352–364, Feb. 1998.

[2] D. G. Jeong and W. S. Jeon, "CDMA/TDD system for mobile multimedia communications," in *Proc. IEEE ICC '98*, Atlanta, GA, June 1998, pp. 994–998. See also, "CDMA/TDD system for wireless multimedia services with traffic unbalance between uplink and downlink," *IEEE J. Select. Areas Commun.*, accepted, 1998.

[3] N. Passas, S. Paskalis, D. Vali, and L. Merakos, "Quality-of-service-oriented medium access control for wireless ATM networks," *IEEE Commun. Mag.*, vol. 35, no. 11, pp. 42–50, Nov. 1997.

[4] ATM Forum Technical Committee, *Traffic Management Specification*, Version 4.0, April 1996.

[5] C. S. Chang, K. C. Chen, M. Y. You, and J. F. Chang, "Guaranteed quality-of-service wireless access to ATM networks," *IEEE J. Select. Areas Commun.*, vol. 15, no. 1, pp. 106–118, Jan. 1997.

[6] M. Shinohara, R. Fan, B. L. Mark, G. Ramamurthy, H. Suzuki, and K. Yamada, "Multicalss large scale ATM switch with QoS guarantee," in *Proc. IEEE ICC '97*, Montreal, Canada, June 1997, pp. 453–461.

[7] M. Schwartz, *Broadband Integrated Networks*, Upper Saddle River, New Jersey: Prentice-Hall, 1996.

[8] R. Guérin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high–speed networks," *IEEE J. Select. Areas Commun.*, vol. 9, no. 7, pp. 968–981, Sept. 1991.

[9] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic theory of a data-handling system with multiple sources," *Bell Syst. Tech. J.*, vol. 61, no. 8, pp. 1871–1894, Oct. 1982.